Gender Pay Gap

1. How many companies are in the data set?

SELECT COUNT(companynumber)

FROM gender_pay_gap_21_22;

There are 10,174 companies in the data set.

2. How many of them submitted their data after the reporting deadline?

SELECT COUNT(submittedafterthedeadline)

FROM gender_pay_gap_21_22

WHERE submittedafterthedeadline = 'true';

There are 361 companies that submitted their data after the reporting deadline.

3. How many companies have not provided a URL?

SELECT COUNT(companylinktogpginfo)

FROM gender_pay_gap_21_22

WHERE companylinktogpginfo = '0'

There are 3,700 companies who did not provide a URL.

4. Which measures of pay gap contain too much missing data, and should not be used in our analysis?

The four largest missing data in the dataset is DiffMeanBonusPercent, DiffMedianBonusPercent, CompanyLinkToGPGInfo, and ResponsiblePerson. This was found by using Excel using the COUNTIF function where the condition would be an empty string. However, DiffMean and DiffMedian might be useful so it was not been removed for further testings.

Bonus (optional): Can you find out what the 'SicCodes' column corresponds to? Is there a way we can understand what each SIC code represents? Search online for extra information.

SicCodes corresponds to the codes that are used to describe the employer's purpose and sectors of work at the time of reporting. This is usually entered by the user manually when adding an employer to their account.

5. Choose which column you will use to calculate the pay gap. Will you use DiffMeanHourlyPercent or DiffMedianHourlyPercent? Can you justify your choice?

SELECT AVG(diffmeanhourlypercent) AS avg_mean,

PERCENTILE_CONT(.50) WITHIN GROUP (ORDER BY diffmeanhourlypercent) AS Median_mean,

AVG(diffmedianhourlypercent) AS avg_median,

PERCENTILE_CONT(.50) WITHIN GROUP (ORDER BY diffmedianhourlypercent) AS Median_median

FROM gender_pay_gap_21_22;

With the query above, I took the mean and the median of the diffmeanhourlypercent and diffmedianhourlypercent and calculated to see the difference between the mean and median of each column.

diffmeanhourlypercent difference (Mean-Median) = 0.735

diffmedianhourlypercent difference (Mean-Meidan) = 2.513

I found out that the difference in the meanhourly is less than the medianhourly, therefore I have decided to use the diffmeanhourlypercent to be used to calculate the pay gap.

6. Use an appropriate metric to find the average gender pay gap across all the companies in the data set. Did you use the mean or the median as your averaging metric? Can you justify your choice?

SELECT AVG(diffmeanhourlypercent) AS avg_mean,

PERCENTILE_CONT(.50) WITHIN GROUP (ORDER BY diffmeanhourlypercent) AS Median_mean

FROM gender_pay_gap_21_22;

I used the average to find the average gender pay gap across all the companies in the data set. Basically the reason I chose to use average is after calculating the difference between the two shown above, I have decided to use a more straightforward and fair number since the difference between the median and mean is only 0.7.

7. What are some caveats we need to be aware of when reporting the figure we've just calculated?

One of the main caveats that I found that we need to be aware of is which one of the diff hourly percent are we using. If we are using the diffmeanhourlypercent, then we will find that the mean and median leans towards equality and balanced as zero equals no gap. However, if we use the diffmedianhuorlypercent, we will find that the gap is leaning towards the male gap.

8. What are the 10 companies with the largest pay gaps skewed towards men?

"HPI UK HOLDING LTD."

"PSJ FABRICATIONS LTD"

"M. ANDERSON CONSTRUCTION LIMITED"

"BIRMINGHAM CITY FOOTBALL CLUB PLC"

"ACUSHNET EUROPE LTD"

"HOOK 2 SISTERS LIMITED"

"CHELSEA FOOTBALL CLUB LIMITED"

"BRAND ENERGY & INFRASTRUCTURE SERVICES UK, LTD."

"MANCHESTER CITY FOOTBALL CLUB LIMITED"

"NEWCASTLE UNITED FOOTBALL COMPANY LIMITED"

9. What do you notice about the results? Are these well-known companies?

I noticed that there are football clubs that in the top 10 which makes sense as most of the employees in those companies are men who are playing the premier league. But also, there are some companies that I haven't heard about which are in the list so for some companies it makes sense as it is focused on male based jobs.

10. Apply some additional filtering to pick out the most significant companies with large pay gaps

SELECT employername, siccodes, diffmeanhourlypercent, diffmedianhourlypercent,

CASE

WHEN employersize = '1000 to 4999' THEN 'n'

WHEN employersize = '5000 to 19,999' THEN 'n'

WHEN employersize = '20,000 or more' THEN 'n'

ELSE 'y'

END AS significant,

employersize

FROM gender_pay_gap_21_22

ORDER by 5, 3 DESC

LIMIT 10;

Female	Male
"FORTEL SERVICES LIMITED"	"CHELSEA FOOTBALL CLUB LIMITED"
"CONCEPT RECRUITMENT GROUP LIMITED"	"BRAND ENERGY & INFRASTRUCTURE SERVICES UK, LTD."
"STAPLES (VEGETABLES) LIMITED"	"NEWCASTLE UNITED FOOTBALL COMPANY LIMITED"
"R E PERSONNEL LIMITED"	
"HYPEROPTIC LTD"	"LEEDS UNITED FOOTBALL CLUB LIMITED"
"REDWIGWAM LIMITED"	"MANCHESTER UNITED FOOTBALL CLUB LIMITED"
"RECRUITMENT SOLUTIONS (SERVICES) LTD"	"THE LIVERPOOL FOOTBALL CLUB AND ATHLETIC GROUNDS LIMITED"
"TIVOLI GROUP LIMITED"	
"ASK ITALIAN RESTAURANTS LIMITED"	"TOTTENHAM HOTSPUR FOOTBALL & ATHLETIC CO. LTD"
"W. H. MALCOLM LIMITED"	"THE ORIGINAL BOWLING COMPANY LIMITED"
	"NORWICH CITY FOOTBALL CLUB PLC"
	"VIRGIN ATLANTIC LIMITED"

Here we filtered a significant company to be a company of a size of 1000 or more employees. Then we selected the top 10 for the male and female which is skewed towards.

11. How would you report on the results? Can we say that these companies are engaging in unlawful pay discrimination?

> The results show that these companies mainly are focused on the specific gender as their workers/employees. For example, most of the male favored payments just have more male workers for example in male football club or construction workers.

12. What's the average pay gap in London versus outside London?

> SELECT COUNT(diffmeanhourlypercent), AVG(diffmeanhourlypercent)
>
> FROM gender_pay_gap_21_22
>
> WHERE address ILIKE '%London%'
>
> The average pay gap in London is 15.68. Which means in average it is more towards the males.

13. What's the average pay gap in London versus Birmingham?

> SELECT COUNT(diffmeanhourlypercent), AVG(diffmeanhourlypercent)
>
> FROM gender_pay_gap_21_22
>
> WHERE address ILIKE '%Birmingham%'
>
> The average pay gap in Birmingham compared to London is 13.21 where as London is 15.68.

14. What is the average pay gap within schools?

> SELECT AVG(diffmeanhourlypercent)
>
> FROM gender_pay_gap_21_22
>
> WHERE siccodes IN ('85100','85200','85310','85320','85410','85421','85422','85510','85520','85530','85590','85600')
>
> The average pay gap within schools is 17.04. This was found by filtering the results to the associated siccodes of schools.

15. What is the average pay gap within banks?

> SELECT COUNT(*), AVG(diffmeanhourlypercent)
>
> FROM gender_pay_gap_21_22
>
> WHERE siccodes IN ('64110','64191')
>
> The average pay gap within banks is 31.27 using the siccodes of banks and financial companies related to banking.

16. Is there a relationship between the number of employees at a company and the average pay gap?

Yes, we have found that whenever a company is small (less than 1000 employees) the average pay gap is more than likely to shift towards one extremity than the other, and mostly towards men.

Bonus (optional): Are there any gender pay gap questions you identified in groups at the start of the session, that we have still not answered using our data set?

Bonus (optional): Take some time to perform any extra analyses, and interrogate the data set with any final questions you have.

Bonus (optional): Once you have answered the 16 questions on the previous slides and submitted your document, you can move on to the next section.