* 1. Gender Pay Gap LAB

1.How many companies are in the data set?

There are 10,174 in the date set.

* SELECT COUNT(employername IS NOT NULL) FROM gender\_pay\_gap\_21\_22;

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

There are 361 companies that have submitted after the deadline.

* SELECT count(\*) FROM gender\_pay\_gap\_21\_22
* WHERE submittedafterthedeadline = 'true' ;

3. How many companies have not provide a URL?

3700 have not provided a URL for pgp information

* SELECT count(\*) FROM gender\_pay\_gap\_21\_22
* WHERE companylinktogpginfo = '0';

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

|  |  |
| --- | --- |
| Measure | Num of zeros |
| SELECT count(\*) FROM gender\_pay\_gap\_21\_22  WHERE diffmeanhourlypercent = '0'; | 99 |
| diffmedianhourlypercent | 861 |
| diffmeanbonuspercent | 2837 |
| Diffmedianbonuspercent | 4019 |
| malebonuspercent | 2704 |
| femalebonuspercent | 2754 |
| malelowerquartile | 229 |
| femalelowerquartile | 218 |
| malelowermiddlequartile | 215 |
| femalelowermiddlequartile | 241 |
| maleuppermiddlequartile | 207 |
| femaleuppermiddlequartile | 248 |
| maletopquartile | 195 |
| femaletopquartile | 242 |

The bonus percentages include a large number of potentially missing data, and thus may not be reliable measures for analysis. There is no way to separate nulls verses true 0 entries.

5. Chose which column you will use to calculate the pay gap. Will you use diffmeanhourlypercent or diffmedianhourlypercent? Justify your choice.

We chose to use the median hourly percent as our metric as we feel that it shows a more accurate representation of the pay gap. This was also the metric chosen by the BBC when they reported the data in previous years.

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

Diffmeanhourlypercent average= 13.64 diffmedianhourlypercent average= 12.31

We used the average of the median because it isn’t affected by outliers so it show more of a true average.

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

For the median there is more missing data, so it may not be fully representative of the population. The mean, on the other hand, can be influenced by outliers on either side of disproportionately positive or negative scores.

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

* SELECT employername, siccodes, diffmeanhourlypercent, diffmedianhourlypercent, address
* FROM gender\_pay\_gap\_21\_22
* ORDER BY diffmeanhourlypercent DESC
* LIMIT 10;

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

A large portion of the top 10 companies are football clubs. These companies are likely to have more male than female employees which may skew the results.

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

The top 5 companies all have less than 500 employees total. The top 3 companies report almost no, or no female employees at all in any of their quartiles.

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

The top companies with little or no female employees would likely not be considered as engaging in unlawful pay discrimination as there is no available comparison. Discriminatory employment practices maybe, but not pay.

12. What’s the average pay gap in London vs outside London?

SELECT \* FROM gender\_pay\_gap\_21\_22 WHERE address LIKE '%London%'

AND address NOT LIKE '%London Road%'

AND address NOT LIKE '%London Street%'

AND address NOT LIKE '%London Luton%'

AND address NOT LIKE '%Londonderry%'

OR postcode = 'EC3R 7LP';

SELECT AVG(diffmeanhourlypercent), AVG(diffmedianhourlypercent) FROM gender\_pay\_gap\_21\_22

WHERE address LIKE '%London%'

AND address NOT LIKE '%London Road%'

AND address NOT LIKE '%London Street%'

AND address NOT LIKE '%London Luton%'

AND address NOT LIKE '%Londonderry%'

OR postcode = 'EC3R 7LP';

* "avg" "avg-2"
* 15.8500000000000000 13.7683720930232558

SELECT \* FROM gender\_pay\_gap\_21\_22

WHERE address LIKE '%London Road%'

AND address LIKE '%London Street%'

AND address LIKE '%London Luton%'

AND address LIKE '%Londonderry%'

OR address NOT LIKE '%London,%'

ORDER BY postcode;

SELECT AVG(diffmeanhourlypercent), AVG(diffmedianhourlypercent) FROM gender\_pay\_gap\_21\_22

WHERE address LIKE '%London Road%'

AND address LIKE '%London Street%'

AND address LIKE '%London Luton%'

AND address LIKE '%Londonderry%'

OR address NOT LIKE '%London,%' ;

"avg" "avg-2"

13.0369429922827981 11.9211227284042818#12 visualization ideas

13. What’s the average pay gap in London vs Birmingham?

SELECT \* FROM gender\_pay\_gap\_21\_22

WHERE address LIKE '%Birmingham%'

AND postcode NOT LIKE 'W%'

ORDER BY postcode;

SELECT AVG(diffmeanhourlypercent), AVG(diffmedianhourlypercent) FROM gender\_pay\_gap\_21\_22

WHERE address LIKE '%Birmingham%'

AND postcode NOT LIKE 'W%';

"avg" "avg-2"

13.2376288659793814 10.9123711340206186

14. What is the average pay gap within schools?

|  |  |
| --- | --- |
| 85100 | Pre-primary education |
| 85200 | Primary education |
| 85310 | General secondary education |
| 85320 | Technical and vocational secondary education |
| 85410 | Post-secondary non-tertiary education |
| 85421 | First-degree level higher education |
| 85422 | Post-graduate level higher education |
| 85510 | Sports and recreation education |
| 85520 | Cultural education |
| 85530 | Driving school activities |
| 85590 | Other education n.e.c. |
| 85600 | Educational support services |

* SELECT AVG(diffmeanhourlypercent) FROM gender\_pay\_gap\_21\_22
* WHERE siccodes LIKE '85%';
  + "avg" 17.3164543524416136
* SELECT AVG(diffmedianhourlypercent) FROM gender\_pay\_gap\_21\_22
* WHERE siccodes LIKE '85%';
  + "avg" 25.7299363057324841

15. What is the average pay gap within banks?

|  |  |
| --- | --- |
| 64191 | Banks |
| 64110 | Central banking |

* SELECT AVG(diffmeanhourlypercent) FROM gender\_pay\_gap\_21\_22
* WHERE siccodes LIKE '64191';
  + "avg" 31.4000000000000000
* SELECT AVG(diffmedianhourlypercent) FROM gender\_pay\_gap\_21\_22
* WHERE siccodes LIKE '64191';
  + "avg" 31.7205882352941176

SELECT AVG(diffmeanhourlypercent) FROM gender\_pay\_gap\_21\_22

WHERE siccodes LIKE '64191%' OR siccodes LIKE '64110%';

"avg" 31.6866666666666667

SELECT AVG(diffmedianhourlypercent) FROM gender\_pay\_gap\_21\_22

WHERE siccodes LIKE '64191%' OR siccodes LIKE '64110%';

"avg"

31.5511111111111111

EXTRA- finding all multi siccode 64191 classifications

SELECT employername,siccodes,diffmeanhourlypercent, diffmedianhourlypercent, address, employersize,responsibleperson

FROM gender\_pay\_gap\_21\_22 WHERE siccodes LIKE '64191%'

ORDER BY diffmeanhourlypercent DESC;

SELECT AVG(diffmedianhourlypercent) FROM gender\_pay\_gap\_21\_22

WHERE siccodes LIKE '64%';

Jus in case(for a laugh)

SELECT \* FROM gender\_pay\_gap\_21\_22

WHERE siccodes LIKE '64%';

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

Larger companies tend to have less pronounced pay gaps than small companies

SELECT employername,siccodes,diffmeanhourlypercent, diffmedianhourlypercent, address, employersize,responsibleperson

FROM gender\_pay\_gap\_21\_22 WHERE employersize = '20,000 or more'

ORDER BY diffmeanhourlypercent DESC

LIMIT 10;

"20,000 or more"

4274 "250 to 499"X

464 "5000 to 19,999"X

2131 "1000 to 4999"X

532 "Less than 250"X

210 "Not Provided" not doing this

2501 "500 to 999"

#16 visualization ideas

* Stacked or cluster chart comparing the middle quartile for men and women at the companies that employ 20,000.
* Maybe small multiples for the rest of the top ten of each company size
* Company names for x axis
* Quartile for the x

**Presentation and visualization ideas and notes**

Under-utilized columns

* Quartile data
* Sic codes

Ideas for these concepts

* Comparing the male to female quartile ratio for biggest and smallest companies with largest pay gap favoring men
* Comparing the male to female quartile ratio for biggest and smallest companies with largest pay gap favoring women
* What sic code group has the largest pay gaps?