Gender Pay Gap Analysis

1- How many companies?

In the original database, there are 10,174 companies.

SQL code:

SELECT COUNT(*)

From gender

Applying my filters, narrowing down to companies with a valid siccode (verified by the UK's siccode database) and a valid postcode (verified by a UK postcode database retrieved online), there are only 6,154 companies. As these values (region and siccode) are important for my analysis, it is crucial that I reduce the number of employers I am analyzing to those with valid/non-NULL data.

SQL code:

SELECT COUNT(*)

FROM gender

JOIN postcodes

ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode

JOIN sic

USING ('siccodes')

2- Late submissions

In the original database, there are 608 companies that submitted their reports after the due date. For this question, I disregard the "SubmittedAfterTheDeadline" column as it contains multiple erroneous values (i.e. rows computed as "False" for employers with a due date prior to their submission date).

SQL code:

SELECT COUNT(*)

FROM gender

WHERE datesubmitted>duedate;

In my filtered dataset, there are 367 companies that submitted their reports after the due date.

SQL code:

SELECT COUNT(*)

FROM gender

JOIN postcodes

ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode

JOIN sic

USING ('siccodes')

WHERE gender.datesubmitted>gender.duedate;

3- Missing URL's

In the original database, there are 3,700 companies with a missing URL, or a URL equal to 0.

```
SQL code:

SELECT COUNT(*)
FROM gender
WHERE companylinktogpginfo='0' or companylinktogpginfo is NULL;
```

In my filtered dataset, there are 2,323 companies with a missing URL, or a URL equal to 0.

```
SQL code:

SELECT COUNT(*)
FROM gender
JOIN postcodes
ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode
JOIN sic
USING ('siccodes')
WHERE gender.companylinktogpginfo='0' or gender.companylinktogpginfo is NULL;
```

4– Missing Data in Columns & Siccodes

The data I downloaded contained 0s, not NULLs. For example, the bonuspercent columns contain many 0s (not NULLS). I am keeping this column in my analysis as 0s do not necessarily signify lack of data.

This goes back to how I filtered my data. As siccodes, representing industry, and postcodes, representing regions/cities, are crucial for this analysis, I will disregard rows that have NULLS/non-valid values in these columns.

Siccodes represent industries. The UK government has a database with an industry description of each code, which I've downloaded for my analysis.

5- Median Pay Gap VS Mean Pay Gap

The median pay gap is the difference between the median pay of a man and the median pay of a woman. The mean pay gap is the difference between the average total earnings of a man and the average total earnings of a woman.

According to the United Kingdom's Office for National Statistics, the median pay gap measure is typically preferred over the mean as it represents the 'typical' employee better. When calculating the mean, results can easily be skewed by few employees paid very high salaries.

Therefore, for better representation, in my analysis I will utilize the median pay gap measure.

6 - Median Pay Gap VS Mean Pay Gap

To obtain the most representative analysis, I will utilize the median per hourly earnings percent to analyze the pay gap. As I work out my analysis, I also refer to the mean per hourly earnings, to analyze any potential significant differences.

7 – Caveats about using Median/Mean Pay Gap

The median earnings measure may not provide the entire picture, as it does not account for outliers, and ignores the distribution of earnings across the dataset.

8– 10 Companies with Largest Pay Gaps Skewed Towards Men

From now on, I refer only to queries for my filtered data (6,154 companies).

The ten companies with the largest pay gaps skewed towards men (representing the highest median difference per hour percentage) are listed below"

ATFC LIMITED
HPI UK HOLDING LTD.
M. ANDERSON CONSTRUCTION LIMITED
PSJ FABRICATIONS LTD
SERVICE INNOVATION GROUP-UK LIMITED
BRAND ENERGY & INFRASTRUCTURE SERVICES UK, LTD.
THE LEARNING FOR LIFE PARTNERSHIP
GREENBROOK HEALTHCARE (HOUNSLOW) LIMITED
MOTORLINE LIMITED
P. D. HOOK (GROUP) LIMITED

SQL code:

```
SELECT *
FROM gender
JOIN postcodes
ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode
JOIN sic
USING ('siccodes')
ORDER BY gender.diffmedianhourlypercent DESC
LIMIT 10;
```

9 – Results Analysis

The companies with the highest median pay gap skewed towards men are all part of different industries, ranging from "primary education" to "manufacture of metal structures" and are all based in different cities across the United Kingdom.

They are not well-known and mainly small: of the ten, 6 employers have less than 499 employees, 2 employers are medium-sized, with 500 to 999 employees, and 1 is medium large with 1000 to 4999 employees, with the last company not having provided the information.

10 – Additional Filtering

Applying some additional filtering based on: employer size and industry.

Filter by Employer Size
The 5 largest (companies with 5000 employees or more) companies with the highest median pay gap skewed towards men are listed below:

EASYJET AIRLINE COMPANY LIMITED
INDEPENDENT VETCARE LIMITED
CVS (UK) LIMITED
H&M HENNES & MAURITZ UK LIMITED
VETPARTNERS PRACTICES LIMITED

The company with the highest median pay gap skewed towards men with 63.6% is Easy Jet, a well-known airline company. The remaining, apart from CVS UK Limited, are in veterinary activities.

SQL code:

```
SELECT *
FROM gender
JOIN postcodes
ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode
JOIN sic
USING ('siccodes')
WHERE employersize IN('5000 to 19,999', '20,000 or more')
ORDER BY gender.diffmedianhourlypercent DESC
LIMIT 5;
```

Filter by Industry

The 3 industries with the highest median pay gap are:
- Repair of other equipment – 53%
- Non-scheduled passenger air transport – 51%

- Retail sale of hearing aids – 44%

However this is not representative as these categories hold only 1-3 companies. Restricting this analysis to industries with 12 or more companies from the dataset (the average number of companies per industry), we find that the following industries have the largest highest median pay gap (skewed towards men):
- Primary Education (112 companies) - 32.26%
- Banks (33 companies) - 31.83%
- Development of building projects (17 companies) - 28.22%

To obtain the average company number per industry:

The 5 companies that have the lowest median pay gap (not skewed towards either men or women) are:

Agents involved in the sale of timber and building materials (16 companies) → -0.59%
Residential care activities for learning difficulties, mental health and substance abuse (26 companies) → -0.00%
Unlicensed restaurants and cafes (128 companies) → 0.25%
Take-away food shops and mobile food stands (36 companies) → 0.75%
Other residential care activities n.e.c. (30 companies) → 0.88%

SQL code:

```
SELECT description, avg(diffmedianhourlypercent),  COUNT(employername)
FROM gender
JOIN postcodes
ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode
JOIN sic
USING ('siccodes')
GROUP BY description
HAVING avg(diffmedianhourlypercent)>-1 and avg(diffmedianhourlypercent)<1 and
COUNT(employername)>12
ORDER BY avg(diffmedianhourlypercent) ASC;
```

11 – Pay Discrimination?

We cannot conclude from this analysis that certain employers engage in unlawful pay discrimination, as pay gaps do not measure unequal pay. Indeed a pay gap does not always signify discrimination.

In 2010, the United Kingdom saw the Equal Pay Act of 1970 superseded for the adoption of the Equality Act. The act includes provisions against discrimination based on age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation. The pay gap can therefore be explained by different reasons, including: fewer women in high paying and senior positions, more women on part-time schedules.

We can conclude, however, that specific industries (banks, primary education schools, and development of project building facilities) tend to have pay gaps skewed towards men, whereas some tend to be less skewed towards either gender (sale of timber and building materials, unlicensed restaurants and takeaway shops, and residential care facilities). These may be places where men and women work similar full-time/part-time hours, with more equitable representation of men and women across seniority levels.

12 – Pay Gap in London VS Outside London

Joining the postcodes table to the gender table, which provides a region column, I can analyze the pay gap difference in London and outside. In the postcodes table, London is broken down into the regions of: City

of London and Greater London. The capital of England has a higher median pay gap than the rest of the United Kingdom (13.11% vs 10.84%).

SQL code:

```
WITH CTE as
(SELECT diffmedianhourlypercent,
CASE WHEN
region LIKE '%London%' THEN 'London'
ELSE 'Other'
END as region_type
FROM gender
JOIN postcodes
ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode
JOIN sic
USING ('siccodes')
WHERE postcodes.region is NOT NULL
)
SELECT region_type, COUNT(region_type) number_companies,
ROUND(AVG(diffmedianhourlypercent),2) as avgdiffmedianhourlypercent
FROM CTE
GROUP by region_type;
```

13 – Pay Gap in Birmingham VS London

Repeating the join of the postcodes table to the gender table, I can analyze the average median pay gap in the region of "Birmingham", which stands at 8.78%, lower than London's at 13.11%. While the pay gap is lower, there are much fewer companies based in Birmingham in this dataset (109 vs 1441 in London).

SQL code:

```
WITH CTE as
(SELECT diffmedianhourlypercent,
CASE WHEN
region LIKE '%London%' THEN 'London'
WHEN region LIKE '%Birmingham%' THEN 'Birmingham'
ELSE 'Other'
END as region_type
FROM gender
JOIN postcodes
ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode
JOIN sic
USING ('siccodes')
WHERE postcodes.region is NOT NULL
)
SELECT region_type, COUNT(region_type) number_companies,
ROUND(AVG(diffmedianhourlypercent),2) as avgdiffmedianhourlypercent
FROM CTE
GROUP by region_type;
```

14 – Pay Gap within Schools

Schools are represented in the description column of the siccodes table under the keyword 'Education'. To ensure I capture any other employer names that may have been mis-labeled, I also include employers with 'school' in the name (I do not include those with 'education' in the name as these can often be services, not schools). There are 454 employers which fall within this filter. The average median pay gap for these employers is 24.26%.

SQL code:

```
SELECT AVG(diffmedianhourlypercent) as avgmedianhourlypercent, COUNT(employername)
FROM gender
JOIN postcodes
ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode
JOIN sic
USING ('siccodes')
WHERE gender.employername
LIKE '%school%'  or sic.Description LIKE '%education%'
```

15 – Pay Gap within Banks

Taking a similar approach to the question above, I consider the keywords that would appear in the siccodes table's Description column for Banks, including: "bank", "banking", "financial services". As some employers across the UK may include the word "bank" in their name, despite not being a bank (such as Southbank Centre), I choose to solely filter by the description column of the siccodes table. This leaves me with 57 companies, the average median pay gap for these is 28.26%.
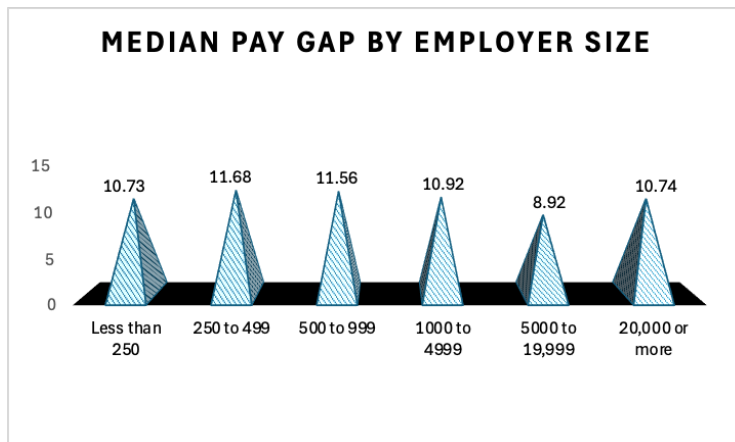
SQL code:

```
SELECT AVG(diffmedianhourlypercent) as avgmedianhourlypercent, COUNT(employername)
FROM gender
JOIN postcodes
ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode
JOIN sic
USING ('siccodes')
WHERE sic.Description LIKE '%bank%' or sic.description LIKE '%financial services%'
```

16 – Relationship between number of employees and the average pay gap

No, there is little relationship between the employee size and the median pay gap, as demonstrated by the chart below. Companies with 5000 to 19,999 employees have the lowest median pay gap, but this category includes fewer companies for comparison (177 vs 2714 for employers with 250 to 499 employees for example).

MEDIAN PAY GAP BY EMPLOYER SIZE

17 – Further Analysis

Analysis into the bonus categories: the male and female bonus categories represent the percentage of male/females which receive promotional bonuses during the recording period. Out of 6,154 companies in the dataset, 2,707 companies have a male bonus percentage higher than the female bonus percentage, 2,030 companies have a female bonus percentage higher than the male bonus percentage, and 1,417 companies have equal bonuses between sexes.

SQL code:

```
SELECT employername, malebonuspercent, femalebonuspercent
FROM gender
JOIN postcodes
ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode
JOIN sic
USING ('siccodes')
WHERE malebonuspercent>femalebonuspercent;


SELECT employername, malebonuspercent, femalebonuspercent
FROM gender
JOIN postcodes
ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode
JOIN sic
USING ('siccodes')
WHERE femalebonuspercent>malebonuspercent;


SELECT employername, malebonuspercent, femalebonuspercent
FROM gender
JOIN postcodes
ON TRIM(substr(gender.postcode,1,INSTR(gender.postcode,' ')))=postcodes.postcode
JOIN sic
USING ('siccodes')
WHERE femalebonuspercent=malebonuspercent;
```

Works Cited

Allegretto, Sylvia A., et al. "The Gender Pay Gap: A Multidimensional Approach." National Bureau of Economic Research, Apr. 2016, https://www.nber.org/papers/w21913. Accessed 17 Nov. 2024.

Equality Act 2010. UK Government, 2010, https://www.legislation.gov.uk/ukpga/2010/15/contents. Accessed 17 Nov. 2024.

"Gender Pay Gap: Six Charts That Explain the Issue." BBC News, 29 Mar. 2019, https://www.bbc.com/news/business-47754005. Accessed 17 Nov. 2024.

"How Covid Has Impacted the Gender Pay Gap." People Management, 10 Mar. 2021, https://www.peoplemanagement.co.uk/article/1745857/how-covid-has-impacted-gender-pay-gap. Accessed 17 Nov. 2024.

Powell, Andrew. The Gender Pay Gap. UK Parliament, House of Commons Library, 28 Feb. 2022, https://researchbriefings.files.parliament.uk/documents/SN07068/SN07068.pdf. Accessed 17 Nov. 2024.

Pidd, Helen. "Gender Pay Gap in Great Britain Smallest Since Reporting First Enforced." The Guardian, 6 Apr. 2024, https://www.theguardian.com/world/2024/apr/06/gender-pay-gap-in-great-britain-smallest-since-reporting-first-enforced. Accessed 17 Nov. 2024.

Radiac. UK Postcodes Dataset. GitHub, https://github.com/radiac/UK-Postcodes/blob/master/postcodes.csv. Accessed 17 Nov. 2024.

"The Enduring Grip of the Gender Pay Gap." Pew Research Center, 1 Mar. 2023, https://www.pewresearch.org/social-trends/2023/03/01/the-enduring-grip-of-the-gender-pay-gap/. Accessed 17 Nov. 2024.

UK Government. SIC 2007: Condensed List of Codes. 2018, https://assets.publishing.service.gov.uk/media/5a7f8639e5274a2e87db65e1/SIC07_CH_condensed_list_en.csv/preview. Accessed 17 Nov. 2024.

"Who Needs to Report." Gender Pay Gap Reporting: Guidance for Employers, UK Government, https://www.gov.uk/government/publications/gender-pay-gap-reporting-guidance-for-employers/who-needs-to-report. Accessed 17 Nov. 2024.