

Project 1: Gender Pay Gap

Table of Contents

1. Problem Statement / Objectives
2. Data Preparation
3. Overview
4. Data Analysis
5. Limitations & Caveats
6. Recommendation on Data Collection Process

1. Problem Statement / Objectives

Problem Statement:

Gender pay differences remain a critical issue in the UK workforce, with variations across industries, regions, and individual companies.

While national policies mandate gender pay gap reporting, the extent of differences and their influencing factors remain complex.

This analysis seeks to uncover trends, industry-specific and regional differences, as well as identifying companies with the most significant pay gaps.

Objectives:

The objective of this analysis is to:

- Determine the national average gender pay gap by analyzing median and mean pay gaps across all reporting companies.
- Compare pay gaps across industries to identify sectors with the largest and smallest differences.
- Analyze regional differences in gender pay differences to determine if certain parts of the UK show higher gaps.
- Identify companies with the largest pay gaps to highlight outliers and potential areas for improvement.

2. Data Preparation

1. Extract

The raw dataset was downloaded from <https://gender-pay-gap.service.gov.uk/viewing/download> and contained 27 columns with a total of 10,564 companies.

Field Name	Data Type	Description
EmployerName	String (Text)	The name of the employer when they reported their data.
EmployerId	Integer	Unique number generated by the gender pay gap service for each employer.
Address	String (Text)	Current registered address the employer.
PostCode	String (Text)	Current registered postcode of the employer.
CompanyNumber	String (Text)	Company number of the employer as listed on Companies House.
SicCodes	String (Text)	The SIC codes that describe the sectors each employer operated in when they reported their data.
DiffMeanHourlyPercent	Float	The mean (average) percentage differences between the hourly pay of men and women.
DiffMedianHourlyPercent	Float	The median percentage differences between the hourly pay of men and women.
DiffMeanBonusPercent	Float	The mean (average) percentage differences between the bonus pay of men and women.
DiffMedianBonusPercent	Float	The median percentage differences between the bonus pay of men and women.
MaleBonusPercent	Float	The percentage of men who were paid a bonus in the reporting year.
FemaleBonusPercent	Float	The percentage of women who were paid a bonus in the reporting year.
MaleLowerQuartile	Float	The percentage of men in the lower hourly pay quarter (lowest paid).
FemaleLowerQuartile	Float	The percentage of women in the lower hourly pay quarter (lowest paid).
MaleLowerMiddleQuartile	Float	The percentage of men in the lower middle hourly pay quarter.
FemaleLowerMiddleQuartile	Float	The percentage of women in the lower middle hourly pay quarter.
MaleUpperMiddleQuartile	Float	The percentage of men in the upper middle hourly pay quarter.
FemaleUpperMiddleQuartile	Float	The percentage of women in the upper middle hourly pay quarter.
MaleTopQuartile	Float	The percentage of men in the upper hourly pay quarter (highest paid).
FemaleTopQuartile	Float	The percentage of women in the upper hourly pay quarter (highest paid).
CompanyLinkToGPGInfo	String (Text)	Link to the 'written statement' published by employers, with more information about their gender pay gap.
ResponsiblePerson	String (Text)	Name of the person confirming that the published gender pay gap information is accurate.
EmployerSize	String (Text)	Number of employees employed by an employer.
CurrentName	String (Text)	The employer's current name.
SubmittedAfterTheDeadline	Boolean	Shows whether employers submitted their gender pay gap data by the relevant deadline ('FALSE') or after the deadline ('TRUE').
DueDate	Date	The deadline date.
DateSubmitted	Date	The actual date that employers submitted their data.

2. Transform

- **Data Preparation:** Duplicates were checked, and no row duplicates were found. Columns that were not relevant for the analysis were removed.
- **Data Cleaning & Standardization:** Missing SIC Codes were filled with "Unknown". Data was extracted for original columns and additional columns were created to enhance the data analysis.

3. Load

The cleaned data consisting of **24 columns** with a total of **10,564 companies** was stored and was accessible for querying, analysis, and visualization with Excel.

Field Name	Data Type	Description
EmployerName	String (Text)	The name of the employer when they reported their data.
EmployerId	Integer	Unique number generated by the gender pay gap service for each employer.
Region	String (Text)	County extracted from address and converted into Region.
Country	String (Text)	County extracted from address and converted into Country.
SICCat1	String (Text)	SICCode converted into SIC Category.
SICCat2	String (Text)	SICCode converted into SIC Category.
SICCat3	String (Text)	SICCode converted into SIC Category.
SICCat4	String (Text)	SICCode converted into SIC Category.
DiffMeanHourlyPercent	Float	The mean (average) percentage differences between the hourly pay of men and women.
DiffMedianHourlyPercent	Float	The median percentage differences between the hourly pay of men and women.
DiffMeanBonusPercent	Float	The mean (average) percentage differences between the bonus pay of men and women.
DiffMedianBonusPercent	Float	The median percentage differences between the bonus pay of men and women.
MaleBonusPercent	Float	The percentage of men who were paid a bonus in the reporting year.
FemaleBonusPercent	Float	The percentage of women who were paid a bonus in the reporting year.
MaleLowerQuartile	Float	The percentage of men in the lower hourly pay quarter (lowest paid).
FemaleLowerQuartile	Float	The percentage of women in the lower hourly pay quarter (lowest paid).
MaleLowerMiddleQuartile	Float	The percentage of men in the lower middle hourly pay quarter.
FemaleLowerMiddleQuartile	Float	The percentage of women in the lower middle hourly pay quarter.
MaleUpperMiddleQuartile	Float	The percentage of men in the upper middle hourly pay quarter.
FemaleUpperMiddleQuartile	Float	The percentage of women in the upper middle hourly pay quarter.
MaleTopQuartile	Float	The percentage of men in the upper hourly pay quarter (highest paid).
FemaleTopQuartile	Float	The percentage of women in the upper hourly pay quarter (highest paid).
EmployerSize	String (Text)	Number of employees employed by an employer.
SubmittedAfterTheDeadline	Boolean	Shows whether employers submitted their gender pay gap data by the relevant deadline ('FALSE') or after the deadline ('TRUE').

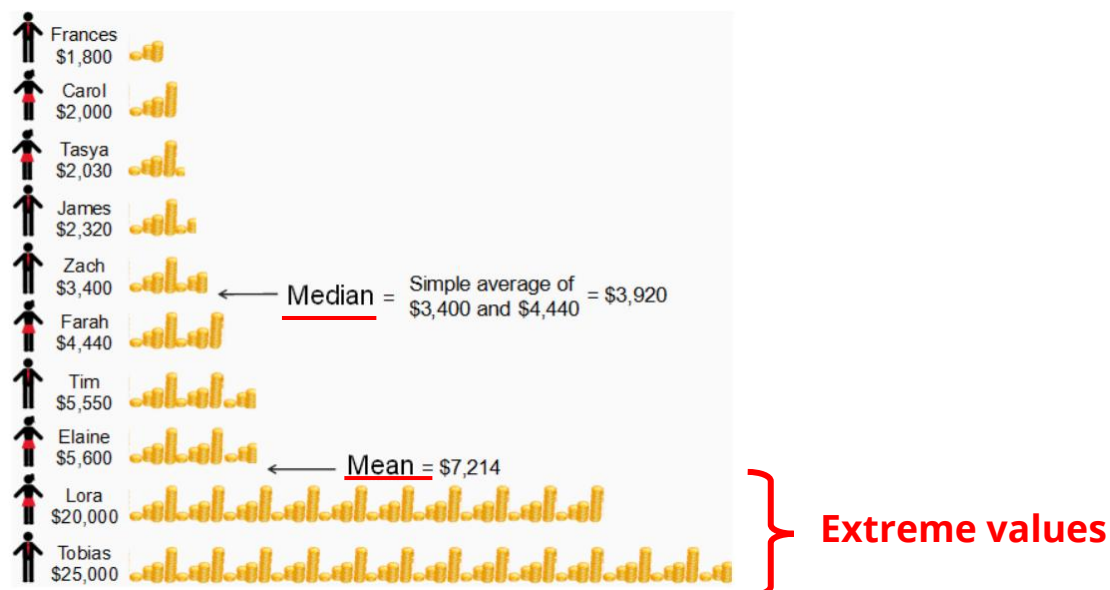
3. Overview

EQUAL PAY ≠ GENDER PAY GAP

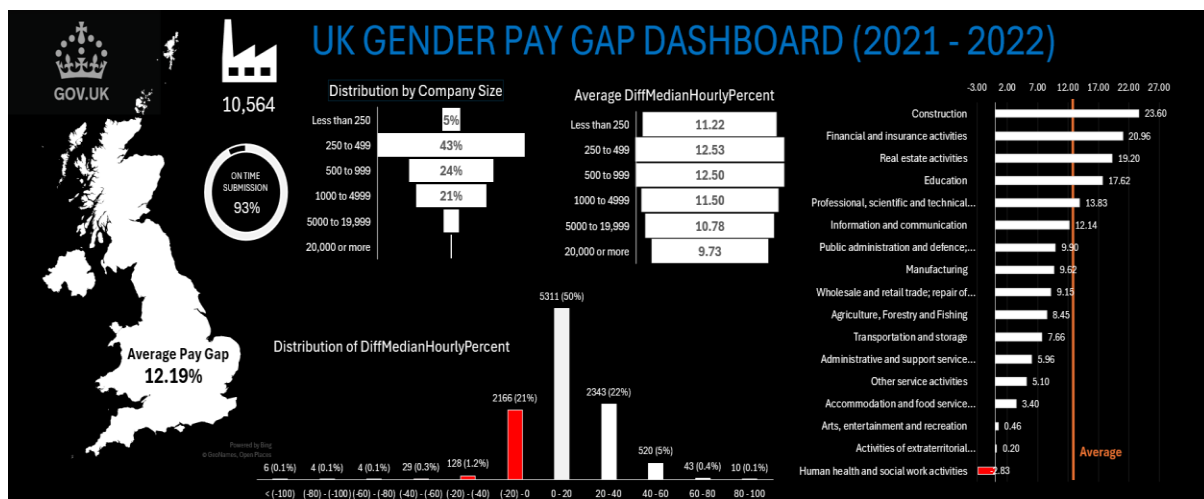
Equal Pay means that men and women performing the same role receive the same pay.

Gender Pay Gap indicates the difference in average earnings between men and women.

To avoid influence of extreme values, median will be used for the analysis of the Gender Pay Gap. Below is an example of how extreme values impact the average.

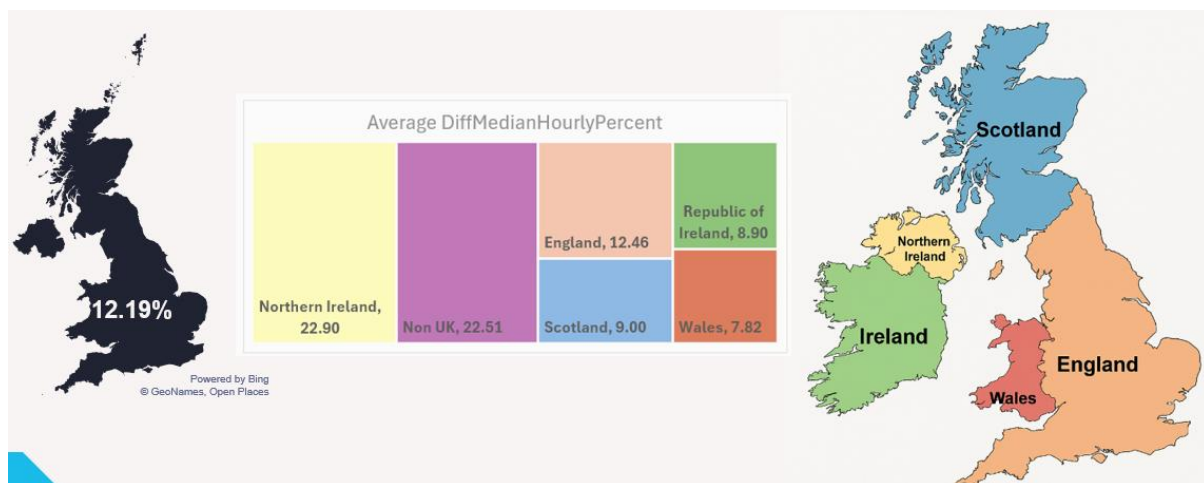


A summarized overview of the analysis can be seen below:



4. Data Analysis

Average Pay Gap for Companies across the UK:



Key Insights:

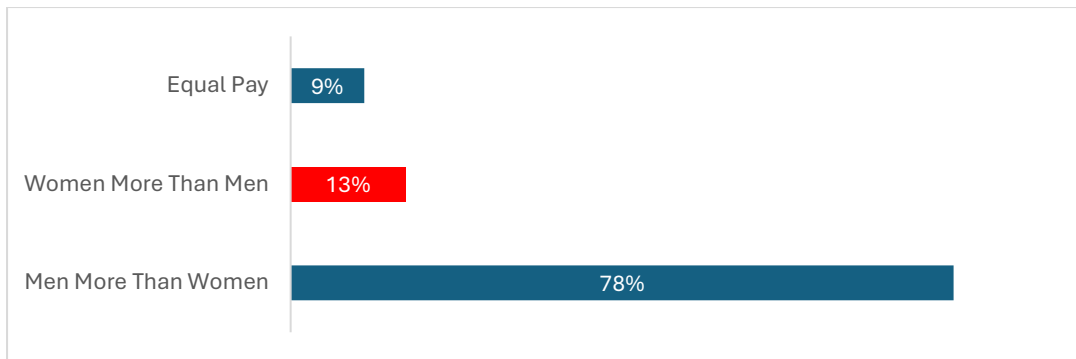
The average median hourly pay gap in the United Kingdom is 12.19%.

Northern Ireland has the highest gender pay gap followed closely by Non UK showing nearly double the UK average.

England is the most aligned with the UK average, showing minimal variance.

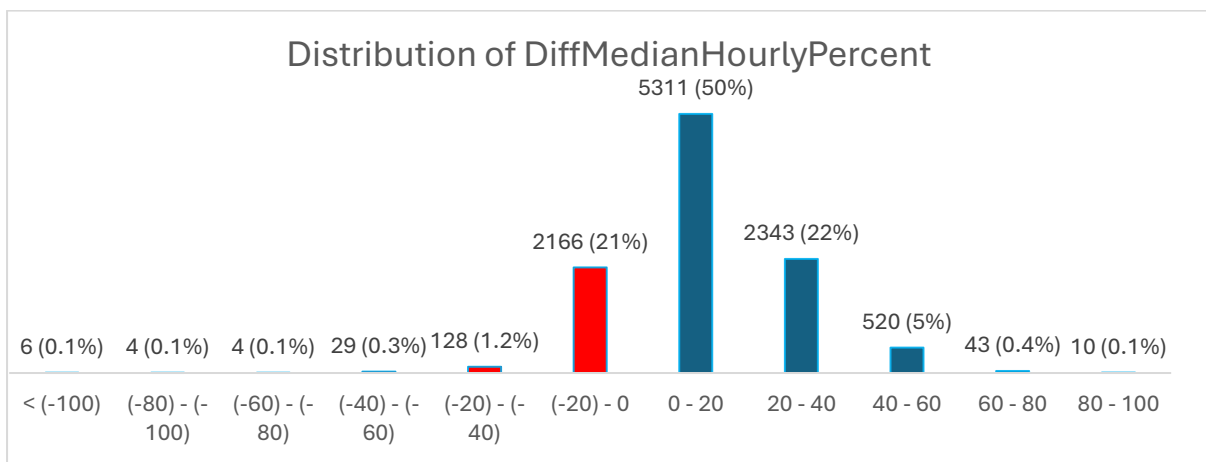
Scotland and the Republic of Ireland have relatively lower pay gaps, indicating more balanced gender pay distributions compared to most UK regions.

Wales shows the lowest gender pay gap.



Key Insights:

The majority of companies (78%) have a median hourly pay gap favoring men, indicating a persistent imbalance. Only 9% have achieved equal pay, and 13% favor women.



Key Insights:

The gender pay gap remains a significant issue, with most organizations skewed towards higher pay for men.

However, a considerable portion (21%) show positive movement towards women, and nearly 1 in 10 companies have achieved equal pay.

The majority of gaps are moderate, but continued effort is needed to address the larger differences and increase the number of organizations with zero gender pay gap.

Differences between the Average Pay Gaps across Industries:



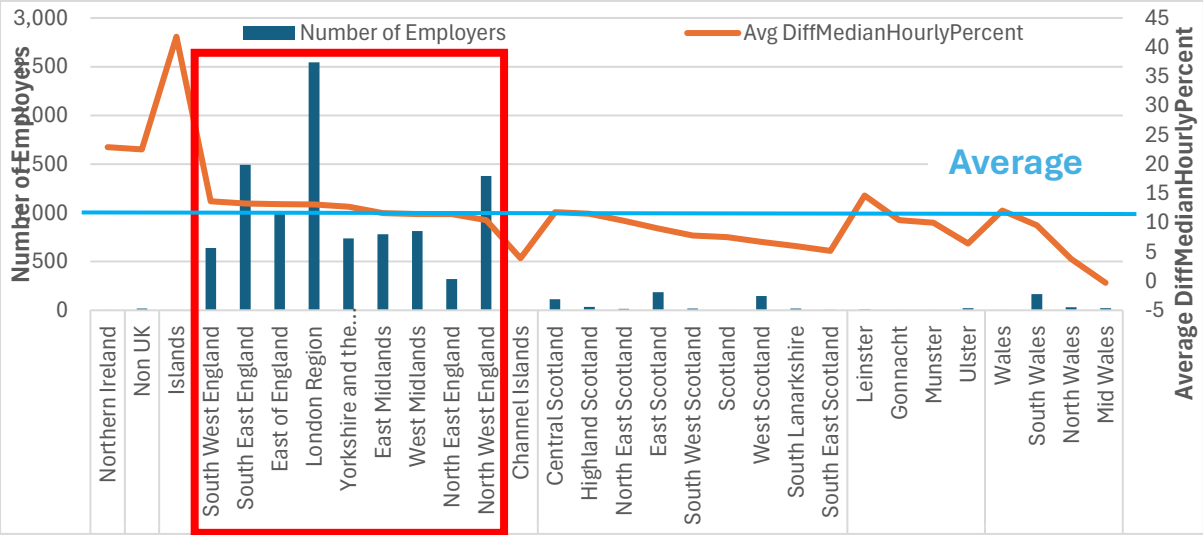
Key Insights:

The top 5 industries show significant pay differences, likely due to seniority imbalances, gender representation in leadership roles, and legacy structures.

The human health and social work sector is the only industry where women earn more than men on average whith arts and service-oriented sectors showing more balnaced pay.

The data reflects broader systemic and cultural factors such as gender representation in leadership roles, industry norms, and occupational segregation.

Differences between the Average Pay Gaps across Different Regions:



Key Insights:

Islands and Non-UK registered companies have extremely high average gaps, though with small sample sizes.

London and surrounding regions (South East, East of England) have higher gaps, possibly due to industry concentration (e.g., finance, tech).

Scotland shows a clear trend of lower pay gaps, especially in western and southern areas.

Mid Wales is the only region with a negative gap, indicating slightly higher pay for women.

Companies with the Largest Pay Gaps:

Top 10 companies who pay men more than women



Top 10 companies who pay women more than men



Key Insights:

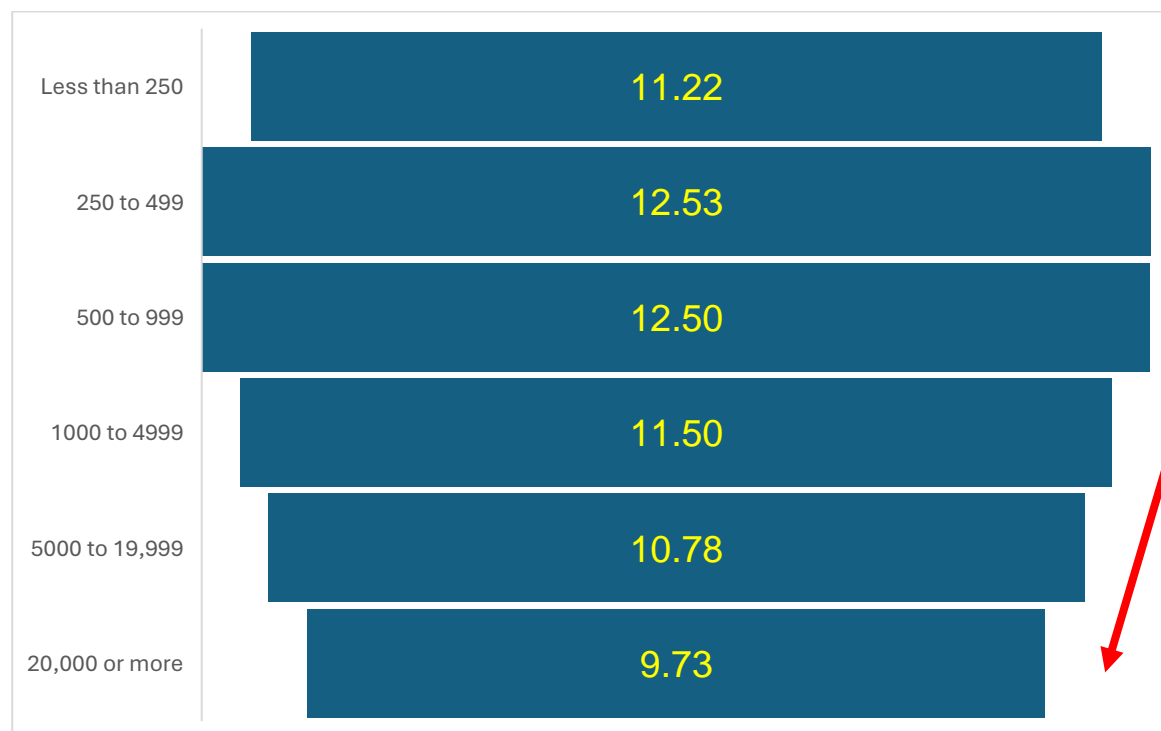
Extremes on both ends indicate that these companies may have:

- Highly gender-skewed workforce distributions
- A lack of gender balance in higher-paid roles
- Potential reporting inconsistencies

Many companies with large male-favored gaps are from construction, infrastructure, or education industries, traditionally dominated by men at higher levels.

Companies with gaps favoring women may operate in care, admin, or payroll outsourcing sectors, where women might dominate higher-paid roles.

Average Pay Gap by Company Size:



Key Insights:

Larger companies (20,000+ employees) have the lowest average pay gap suggesting:

- Greater investment in equality initiatives
- More formalized HR / pay structures
- Greater visibility and accountability

Company size appears to influence gender pay gap, with larger organizations performing better on average.

The data suggests a potential benefit of larger scale companies who have more resources and policies to address gender-based pay differences.

Smaller and mid-sized organizations may need more support or guidance in closing gender pay gaps.

5. Limitations & Caveats

Following are the limitations and caveats that should be considered when interpreting above findings:

1. Gender Pay Gap ≠ Equal Pay

- The gender pay gap measures average earnings differences between men and women across all roles, but it does not indicate whether men and women are being paid equally for the same job.
- A company could have a gender pay gap due to fewer women in senior roles, even if they have an equal pay policy.

2. Lack of Industry-Specific Context

- Some industries (Construction, Finance, and Technology) have historically fewer women in leadership positions, which inflates the pay gap.
- Comparisons across industries should account for structural differences, such as workforce composition and role distribution.

3. No Adjustments for Part-Time & Flexible Work

- The dataset does not account for differences in working hours, such as the fact that women are more likely to work part-time, which affects average pay.
- A high gender pay gap may partly reflect work-life balance choices, rather than outright pay discrimination.

4. No Consideration of Experience & Qualifications

- The dataset does not include information on education, experience, or job tenure, which are key factors in determining pay.
- A reported pay gap may not necessarily be due to bias but could be influenced by differences in career progression.

5. Data Quality & Compliance Issues

- Some companies may report incorrect or incomplete data, and smaller firms (with fewer than 250 employees) are not required to report at all.

- We need to be cautious when analyzing companies with missing, inconsistent, or outlier data.

6. No Breakdown by Job Role or Seniority

- The dataset does not show pay gaps within specific job roles or levels (e.g., entry-level vs. executive roles).
- Large pay gaps might indicate underrepresentation of women in higher-paying roles, rather than direct pay discrimination..

6. Recommendation on Data Collection Process

Improving data collection through validation, completeness checks will ensure more accurate, reliable, and insightful gender pay gap reporting across the UK.

1. Standardization

- Introduce a standardized format for employer names, addresses, and SIC codes to improve categorization.

2. Improve Data Completeness

- Ensure all companies provide SIC codes to allow for accurate industry analysis. (7% of missing SIC codes)
- Ensure that key mandatory fields are provided in the correct format to allow for proper analysis.
- Define clearly the roles of the ResponsiblePerson to improve accountability on data accuracy.

3. Workforce Composition & Role-Based Data

- Provide a breakdown by employment type (Full-time vs. Part-time).
- Job Role Categories – Breakdown of employees by seniority levels (e.g., entry-level, mid-management, senior leadership)