Exploratory Data Analysis Report - Ireland

Dataset: Irish_GPG_2022_2024_cleaned.csv

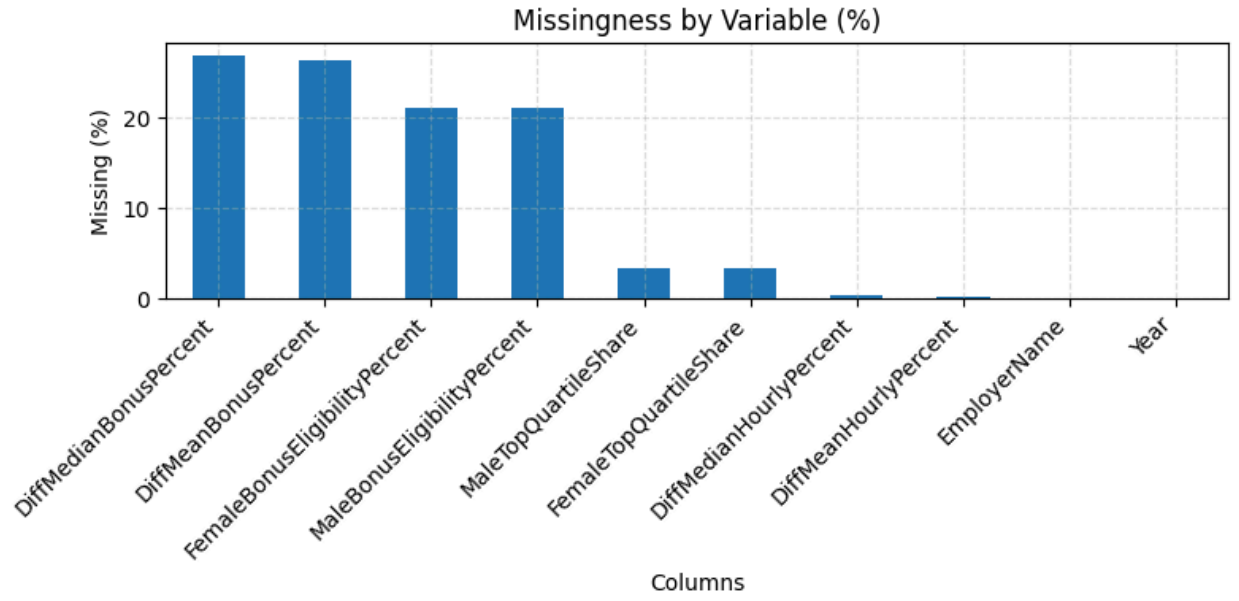
Introduction

The dataset contains employer-reported statistics for Ireland gender pay gap filings. This report focuses on data quality, distributions, temporal trends, and relationships among reported metrics. To assess the quality and structure of Irish_GPG_2022_2024_cleaned.csv and extract decision-ready insights by:

- **Profiling data coverage and schema** – validating column types, record counts per year, and high-level structure of the employer-year panel.
- **Quantifying missingness and reporting gaps** – summarising missing values by variable and year, and identifying which metrics are structurally under-reported (e.g., bonus fields).
- **Characterising distributions of key metrics** – examining the shape and spread of hourly and bonus pay gaps, with and without outliers, for the latest year.
- **Analysing temporal trends (2022–2024)** – tracking changes in mean hourly and bonus pay gaps, female top-quartile representation, and bonus eligibility over time.
- **Assessing representation and reward structures** – comparing male vs female bonus eligibility and examining female share in the top pay quartile.
- **Investigating relationships between metrics** – using scatterplots and correlation matrices to understand how pay gaps, bonus gaps, and representation metrics move together.
- **Highlighting risk areas and improvement opportunities** – distilling the EDA into actionable findings for policymakers and employers (e.g., widening bonus gaps, stalled progress in senior representation).

1. Overview of Missingness by Variable

Understanding missing values is crucial to evaluating data quality and ensuring that downstream analysis is reliable.



The bar chart below displays the percentage of missing values for each variable.

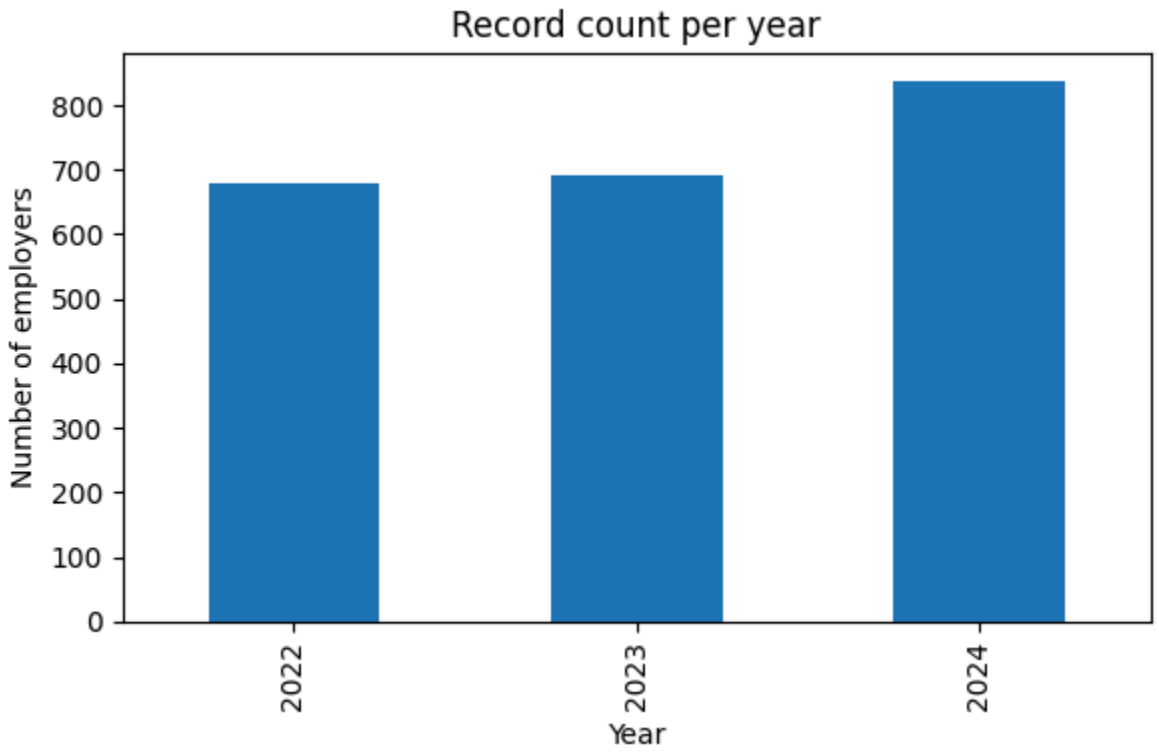
Key Findings:

- **Bonus-related fields** have the highest missingness:
  - DiffMedianBonusPercent: **27%**
  - DiffMeanBonusPercent: **26%**
  - FemaleBonusEligibilityPercent: **21%**
  - MaleBonusEligibilityPercent: **21%**
- **Top quartile representation fields** show low but noticeable missingness:
  - FemaleTopQuartileShare: **3–4%**
  - MaleTopQuartileShare: **2–4%**
- **Hourly pay gap variables** have **almost no missingness** (<1%).
- EmployerName and Year have **0% missing**, meaning each record is fully identifiable and associated with a reporting year.

Interpretation:

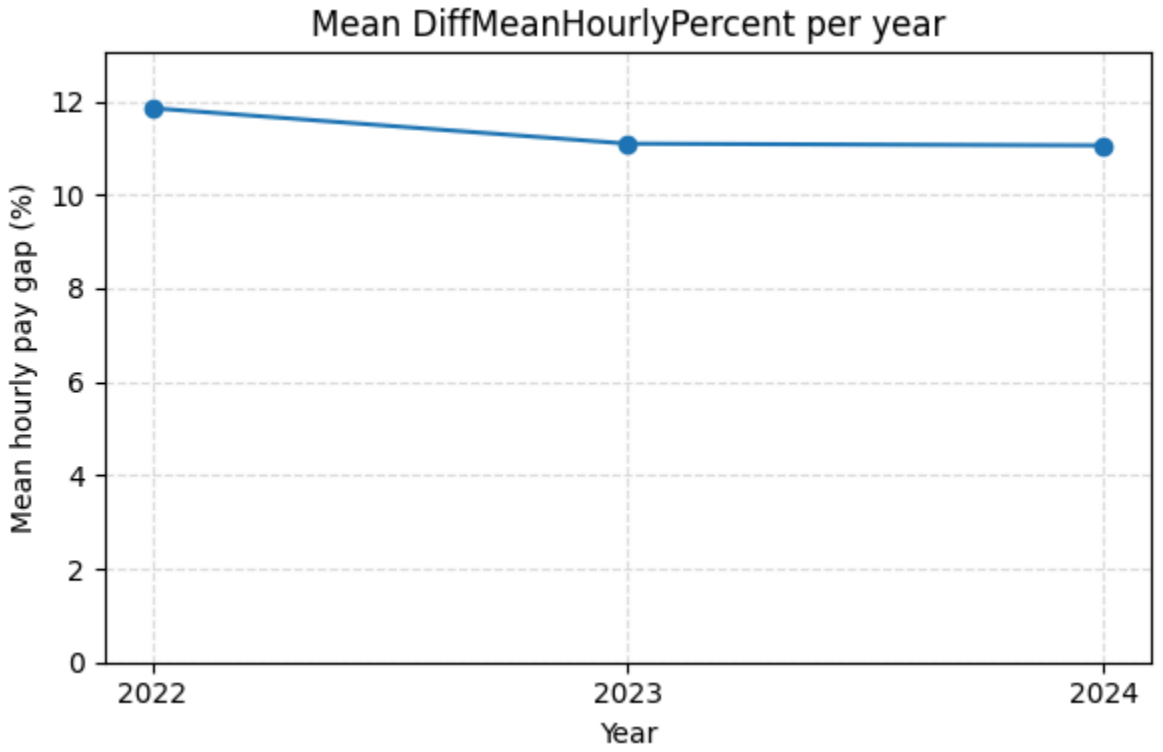
- Bonus metrics are the most incomplete because **not all employers offer bonuses**, and some firms do not report bonus data consistently.
- Top quartile representation is mostly complete but slightly missing for employers that **do not disclose quartile breakdowns** or have incomplete returns.
- Hourly pay fields are nearly always present because they are **mandatory disclosures** for reporting employers.

2. Record Count per Year



The number of employers reporting Gender Pay Gap data increased steadily from 2022 to 2024, rising from roughly 680 to over 840 submissions. This growth likely reflects expanded reporting compliance and improved data ingestion rather than a genuine change in the number of eligible employers. As coverage improves, the dataset becomes more representative and reliable, especially in 2024, which provides the strongest basis for year-on-year comparisons.

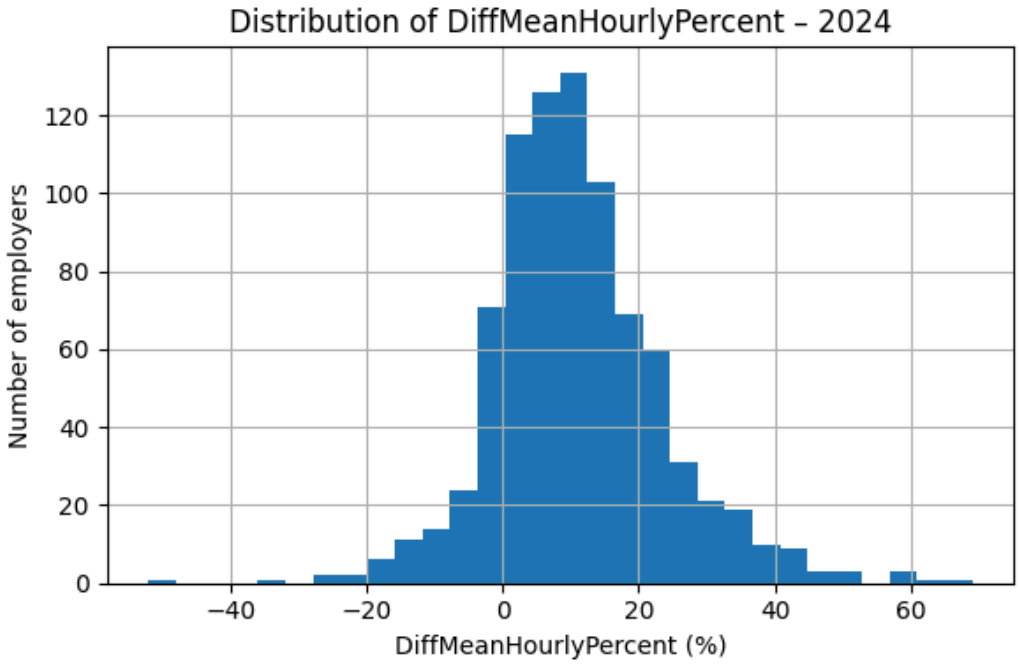
3. Mean Hourly Pay Gap (DiffMeanHourlyPercent)



The mean hourly pay gap shows the percentage difference between men's and women's average hourly earnings, indicating how much less (or more) women are paid per hour.

The average mean hourly gender pay gap shows a **gradual improvement** from **2022 to 2024**. It falls from just under **12% in 2022** to about **11% in 2023**, and remains relatively stable in **2024**. This slow but consistent decline suggests **modest progress in hourly wage equality**, though the gap remains significant, indicating that structural differences in pay persist across employers.

3. Distribution of DiffMeanHourlyPercent



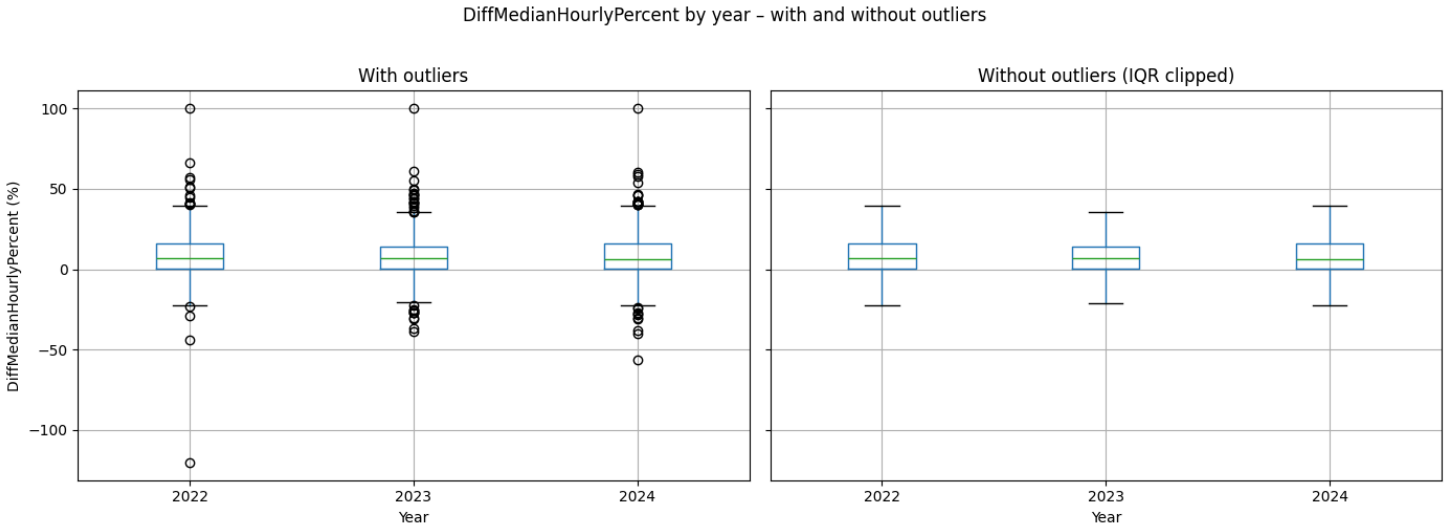
This histogram shows the distribution of mean hourly gender pay gaps across all employers in 2024, essentially, how common different pay gap levels are.

- Most employers have a **positive pay gap**, meaning men earn more per hour on average.
- The distribution peaks around **10–15%**, showing that this is the most common gap range.
- Some employers report **very small gaps**, close to **0%**, indicating near parity.
- A few employers report **negative gaps**, where women earn more.
- There are also **extreme outliers** on both sides, though they represent very few employers.

In 2024, the typical employer pays women about 10–15% less per hour than men. The majority of firms cluster around this central range, with only a small proportion achieving parity or showing reversed gaps. This distribution suggests:

- The gender pay gap is **widespread**, not limited to a few employers.
- Most firms fall within a similar band, indicating systemic patterns rather than isolated issues.
- Outliers likely reflect **industry-level differences**, **small sample effects**, or **unique workforce structures**.

4. DiffMedianHourlyPercent by Year - With vs. Without Outliers



These boxplots compare the **median hourly gender pay gap** across employers for 2022–2024, first including all reported values, and then with extreme outliers removed using an interquartile range (IQR) filter.

1. Median pay gaps are consistently positive

Across all three years, the median employer reports a gap of around **8–12%**, meaning women earn **8–12% less** than men at the midpoint of the pay distribution.

2. Outliers distort the raw distributions

With outliers included, the plots show:

- Very large **positive gaps** (some >50% where men earn far more)
- A smaller set of **negative gaps** (where women earn more)

These extreme values inflate the whiskers and make the central pattern harder to see.

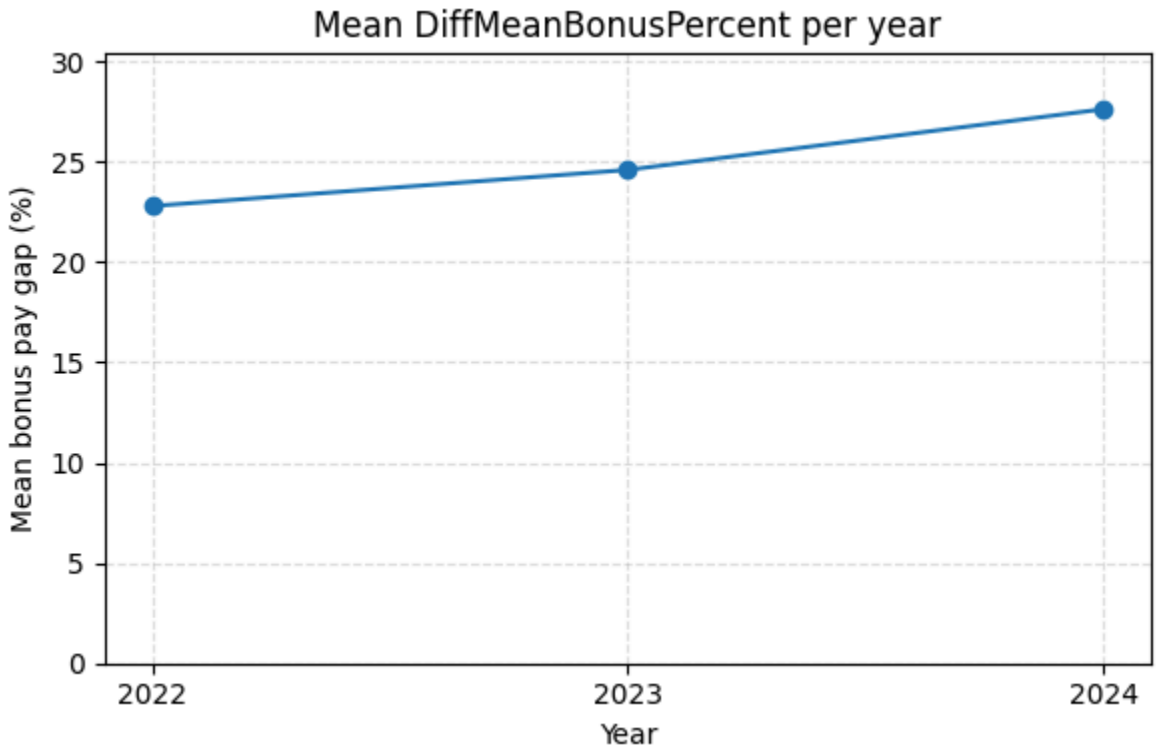
3. After removing outliers, the underlying pattern becomes clearer

The IQR-clipped boxplots reveal:

- A **tight, stable distribution** across years
- Typical gaps consistently falling in the **5–15%** band
- Slight upward movement in 2024 but no major structural shift

Even after removing extreme reporting values, the median hourly gender pay gap remains persistent and stable, indicating systemic wage inequality rather than year-specific anomalies.

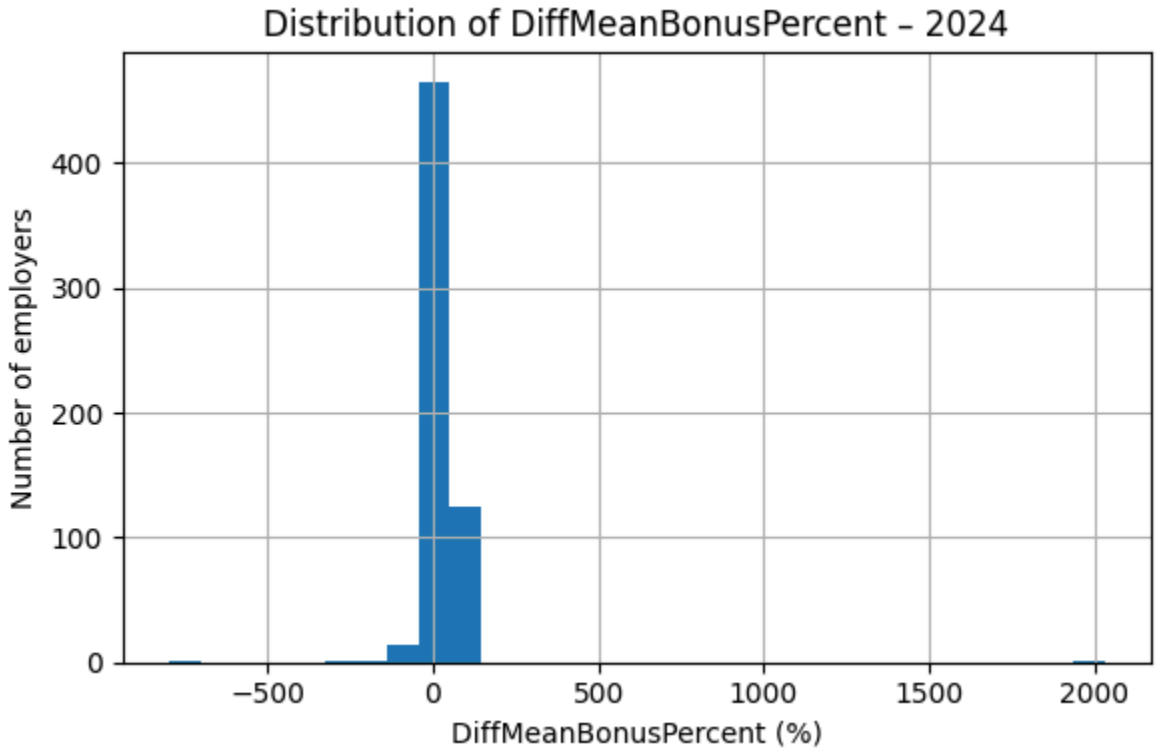
5. Mean Bonus Pay Gap (DiffMeanBonusPercent)



This chart tracks the average mean bonus pay gap across employers from 2022 to 2024, measuring how much less (or more) women receive in bonus pay compared to men.

- The mean bonus gap is **significantly higher** than the hourly pay gap, rising from **~23% in 2022** to **~28% in 2024**.
- This upward trend suggests that **bonus structures are becoming increasingly unequal**, even as hourly pay gaps remain more stable.
- The widening gap may reflect factors such as:
    - Bonuses tied to **senior or commission-based roles** where women are under-represented
    - **Performance-linked awards** that favor roles held predominantly by men
    - **Discretionary bonus practices**, which tend to create more variation and inequality

6. Distribution of Mean Bonus Pay Gap (2024)



This histogram displays how the mean bonus pay gap varies across employers in 2024 i.e., the percentage difference between men's and women's *average* bonus payments.

1. The vast majority of employers cluster between 0% and ~150%

Most firms report bonus gaps in the 0–150% range, meaning men typically receive significantly higher bonuses on average.

2. Extreme outliers exist on both ends

- A small number of employers show very large positive gaps, including values above 1000–2000%, where men's bonuses far exceed women's.
- A few employers show negative gaps (e.g., –200% to –600%), where women receive higher bonuses.

These extreme cases likely reflect:

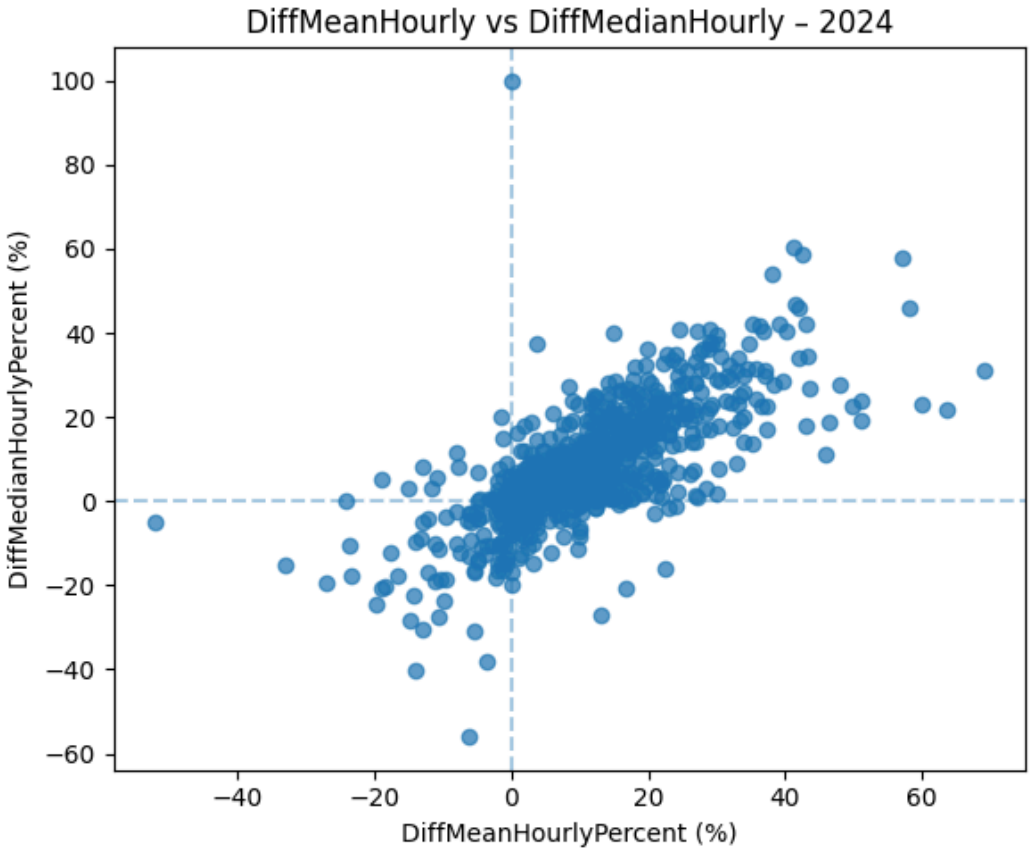
- Small employee populations
- Highly skewed commission roles
- Executives receiving very large discretionary bonuses
- Data entry/reporting anomalies

3. The distribution is heavily right-skewed

The long right tail indicates that a handful of firms drive the overall average up, which explains why mean bonus gaps appear so large in earlier charts.

Bonus pay gaps vary widely across employers and are heavily skewed by extreme outliers, but the central pattern shows that men receive substantially higher bonuses than women in most organizations.

7. Mean vs. Median Hourly Pay Gaps (2024)



This chart compares each employer's **mean hourly pay gap** with their **median hourly pay gap**, showing how the two measures align.

i. Strong positive relationship

Most points fall along an upward diagonal, indicating that employers with a **high mean pay gap** also tend to have a **high median pay gap**. This suggests that gender pay disparities are consistent across the distribution, not just driven by a few high earners.

ii. Clustering around positive gaps

The dense cluster in the **0–20% range** on both axes shows that: Most employers pay women **10–20% less per hour** on both mean and median measures.

iii. Outliers exist but are not dominant

Some employers appear far from the cluster:

- **Far-right outliers** → large gaps favoring men
- **Far-left or negative outliers** → gaps favoring women

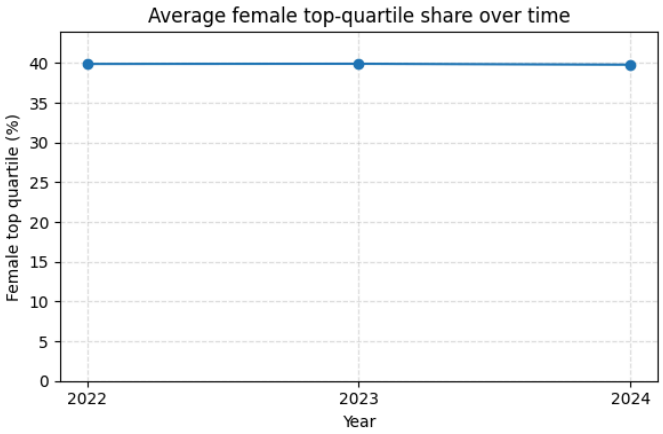
These cases may reflect specific workforce structures, small sample groups, or reporting anomalies.

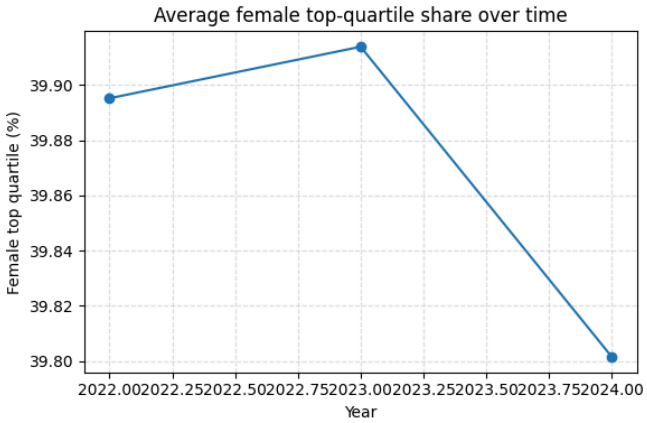
iv. Mean gaps tend to be larger than median gaps

The points generally lie **above the diagonal**, meaning: Pay gaps sometimes widen when higher-paid roles or bonus elements disproportionately benefit men.

Employers with large gender pay gaps typically show them consistently across both mean and median measures, confirming that the gap is broad-based rather than driven by a few extreme salaries.

8. Average Female Top-Quartile Share Over Time





Female Top-Quartile Share shows the percentage of the highest-paid roles (top 25% of earners) that are held by women.

Key insight:
Across 2022–2024, the share of women in top-quartile roles remains **flat at ~40%**, showing **no meaningful improvement in senior-level gender representation**.

Detailed reading:

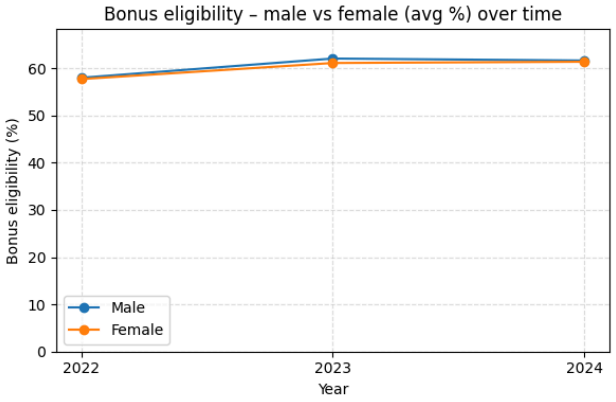
- **2022:** ~39.9%
- **2023:** slight uptick to ~39.92%
- **2024:** drop to ~39.8%

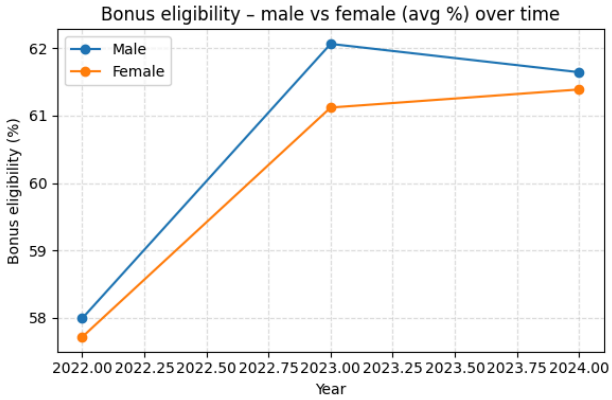
This fluctuation is **statistically negligible**, indicating:

- No major gains in women entering top-paying roles
- Persistent structural barriers in senior leadership progression
- Employers may be improving gender balance in lower/middle levels but **not yet translating this into leadership equality**

Despite other improvements, representation of women in the top quartile of earners has remained effectively unchanged at ~40% over the last three years, signaling ongoing challenges in advancing women into senior, high-paying roles.

9. Interpretation of Bonus Eligibility Over Time (2022–2024)





Both men and women have **similar bonus eligibility rates**, and the gap between them is **very small**.

Key findings

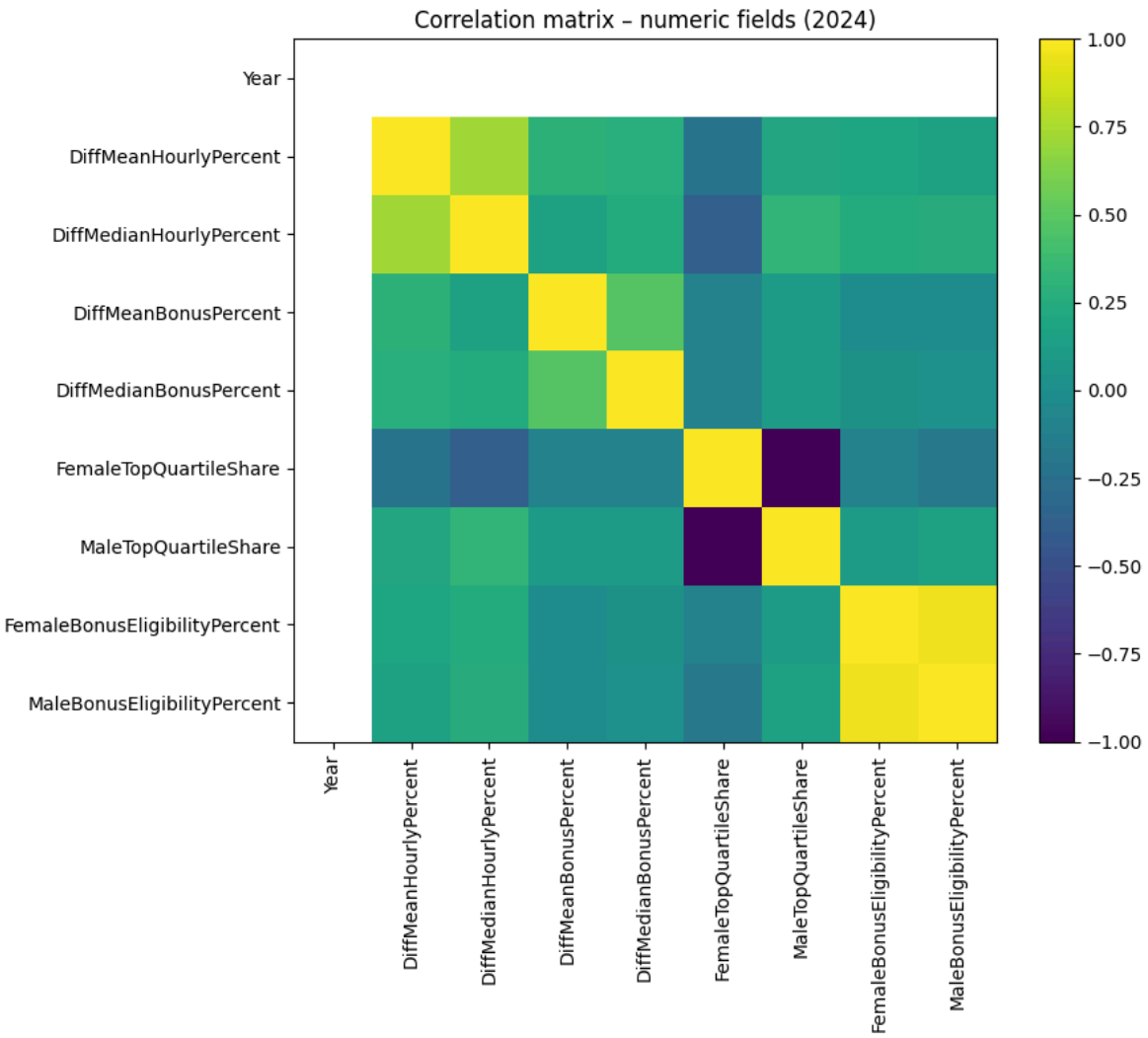
- **Eligibility increased for both groups** from 2022 → 2023, then leveled off in 2024.
- Men consistently remain **1–2 percentage points higher** than women — a **small but persistent gap**.
- **Both genders hover slightly above 60% eligibility** by 2024.

Bonus eligibility rates are high and improving for both men and women, but a small and persistent gap (1–2 percentage points) remains, with men slightly more likely to receive bonuses each year.

What this might signal

- The organisation is **fairly consistent** in providing bonus opportunities across genders.
- However, the slight male advantage may reflect differences in:
    - eligibility-linked job roles
    - seniority levels
    - performance thresholds
    - bonus policies that indirectly favour certain job families

10. Correlation Matrix Interpretation (2024)



The chart displays how strongly different gender-pay-gap metrics move together. Values close to

+1 = strong positive relationship,
−1 = strong negative relationship,
0 = no relationship.

Key Insights

Mean vs Median Pay Gaps → Strong Positive Correlation

- DiffMeanHourlyPercent ↔ DiffMedianHourlyPercent
- DiffMeanBonusPercent ↔ DiffMedianBonusPercent

Employers who have a high mean pay gap also tend to have a high median pay gap; the gaps are structural, not caused by a handful of extreme salaries.

Hourly Pay Gaps show a weak-to-moderate link with Bonus Gaps

- Correlation is positive but not strong.

Hourly pay inequality and bonus inequality are related but not driven by the same factors.

This suggests different mechanisms:

- Base pay differences (role, grade, hours)
- Bonus differences (eligibility rules, performance criteria)

Female vs Male Top-Quartile Share → Strong Negative Correlation

- FemaleTopQuartileShare is strongly negatively correlated with MaleTopQuartileShare (≈ −1).

This is expected, the more women in senior/top roles, the fewer men, and vice-versa. It confirms the dataset is internally consistent.

Bonus Eligibility (Male & Female) are Highly Correlated

- FemaleBonusEligibilityPercent ↔ MaleBonusEligibilityPercent ≈ 0.9+

Companies tend to have similar bonus eligibility rates for both genders (i.e., eligibility is a company-wide policy, not gender-specific). The small, persistent gap shows up in averages, not correlation.

Top-Quartile Share has Weak Relationship With Pay Gaps

Moderate negative correlation between:

- FemaleTopQuartileShare ↔ (pay gap metrics)

Companies with more women in senior roles tend to have **lower gender pay gaps**, but the relationship is not extremely strong, senior representation helps, but doesn't fully explain the gap.

Use of AI:

Used Gen AI to debug the code and draw in depth insights from the EDA.