Data Request for Pay Gap Analysis

This document specifies the HR data required to perform EU-compliant gender pay gap analysis under Directive 2023/970.

Required Data Format

One record per employee with the following fields:

Field	Type	Description	Example
employee_id	string	Unique identifier (anonymised)	"EMP001"
gender	string	"Female" or "Male"	"Female"
annual_salary_eur	integer	Base annual salary in EUR (excluding bonuses)	65000
pay_grade	string	Job classification level	"E", "F", "G"
age	integer	Employee age in years	34
tenure_years	float	Total years at company	5.5
tenure_grade_years	float	Years in current pay grade	2.0
performance_rating	float	Performance score (1-5 scale, 3-year average preferred)	3.8

Data Volume

- **Sweden**: ~750+ records (confirmed), ~29% female
- **Worldwide**: 5,000+ records total

Known Data Issues

- **Performance rating**: may be missing for Sales organisation - can we get sales targets met (% of quota) as alternative metric? How many employees are in Sales? We can build a surrogate model if needed

- **Tenure**: request from Valerie & Jolanta - preference to consider tenure in current pay grade rather than total tenure. **Request**: please provide both total tenure and grade tenure so we can evaluate which has stronger predictive power

- **Base salary reference**: ideally provide the reference "base salary" that other salaries are multiplicative of - this helps understand the pay structure and grade multipliers

- **Payrise data** (optional): as additional proxy for performance, can we get recent payrise broken down into: (1) last payrise at current pay grade, (2) last payrise across pay grade (due to promotion). Not required - just interesting to see the effects

---

Delivery Format

- **Format**: Excel (.xlsx)
- **File naming**: `hr_paygap_export_YYYY-MM-DD.xlsx`

FAQ

Q: What if someone was just hired - say started a month ago? Does that count the same as someone hired at the end of the year?

A: Yes, if we count years not months for tenure. With large sample sizes, the positive and negative differences versus beginning of year cancel out.

Q: What if someone on a high grade was just hired? Do we count their prior experience?

A: No. We assume their grade coefficient is a compound one that implicitly factors in experience and other attributes with respect to the outcome.

Q: What is the J Grade coefficient composed of?

A: J Grade coefficient = Intrinsic Experience coefficient + Knowledge coefficient + other factors. Document the weighting formula if available.

Q: Would the model be better if we consider current position tenure vs total tenure?

A: The model will tell. We calculate coefficients and their respective standard errors for both variables. If statistical significance is below 0.05, we can discard them as individual contributions and instead combine them into one larger factor (total tenure vs discrete tenure with respect to position).

Q: What if we do not have complete performance data for some persons - say freshly hired or not reviewed?

A: We can handle this in three ways:
1. **Exclusion**: the model simply excludes these individuals from performance-related calculations
2. **Mean imputation**: replace missing values with the mean for that pay grade - preserves sample size and point estimates remain stable, but may slightly underestimate uncertainty
3. **Stochastic imputation**: sample from the distribution of performance ratings within that pay grade - this preserves both sample size and variance, giving more accurate uncertainty estimates

For small proportions of missing data, all approaches yield similar results.

Questions? Contact Konrad Jelen at konrad.jelenext@delaval.com