# Dataset Overview: People Analytics Simulation

## 1. hrsystem\_data.csv

Contains employee record data that is usually found in hr systems that track employment history. This dataset only holds the current position of each employee (or last position if they have left the company) from 2000 until the end of 2024.

### Columns

|  |  |
| --- | --- |
| start\_date | The exact date when an employee joined the company |
| department | The department an employee works in (e.g. Sales, HR etc.) |
| job\_level | One of 5 job levels that define seniority in the organization (e.g. entry level) |
| country | One of the locations where the employees are based |
| leaving\_date | The date when an employee left the company, otherwise blank |
| employee\_status | Everyone who joined before the end of 2024 and has not left until the end of 2024 is defined as a current employee |
| employee\_id | Unique identifier for every employee |
| start\_year | The year when an employee joined the company |
| start\_month | The month of a particular year when an employee joined the company |
| leaving\_year | The year when an employee left the company |
| leaving\_month | The month of a particular year when an employee left the company |
| tenure\_years | Number of years an employee was employed at the company in total when they left or their current tenure until the end of 2024 |
| update\_status | This is just a column that indicates that this is the current profile of an employee (or the last profile of an employee who left). This is not relevant for this analysis but would be relevant if we want to look at past job history |
| gender | Gender (Male, Female, Other) |
| performance\_24 | Performance rating for 2024, only current employees received a performance rating at the end of the year |

## 2. applicant\_data.csv

Tracks each applicant's progress through the four stages of the recruitment funnel (CV screening to final interview), including status ("rejected", "progressed", or "hired") and anonymised IDs for job and applicant.

### Columns

|  |  |
| --- | --- |
| employee\_id | Unique identifier for every employee |
| job\_level | One of 5 job levels that define seniority in the organization (e.g. entry level) |
| department | The department an employee works in (e.g. Sales, HR etc.) |
| start\_month | The month of a particular year when an employee joined the company |
| status | The status assigned to every applicant at every assessment stage (hired, progressed or rejected) |
| stage | The application stage and the name of the application stage (e.g. stage 1: CV Screening comes before Stage 2: Telephone Interview |
| applicant\_id | Unique identifier for every applicant |
| job\_id | Unique identifier for every job that people applied and interviewed for |
| gender | Gender (Male, Female, Other) |
| question\_1 | User-defined interview question. |
| question\_2 | User-defined interview question. |
| question\_3 | User-defined interview question. |

## 3. feedback.csv

360-degree feedback scores for current leadership roles with employee ID and with ratings from self, manager, peers, and direct reports across 12 leadership competencies. Columns are labelled in the format RaterType: Competency. Find a description of the competencies below.

|  |
| --- |
| **Inspire with Purpose** |
| *Clearly communicates a compelling vision and gives meaning to goals and work.* |
| **Drive for Results** |
| *Maintains focus on goals and holds self and others to high standards of performance.* |
| **Own the Outcome** |
| *Takes responsibility for results and follows through on commitments.* |
| **Speak Up and Listen** |
| *Encourages open dialogue and actively listens to different viewpoints.* |
| **Lead with Empathy** |
| *Shows understanding for others' perspectives and supports emotional wellbeing.* |
| **Grow Others** |
| *Develops talent through feedback, coaching, and creating learning opportunities.* |
|  |
| **Bridge Across Boundaries** |
| *Collaborates across teams and departments to achieve shared goals.* |
|  |
| **Act with Integrity** |
| *Models ethical behaviour, treats people fairly, and stays consistent under pressure.* |
|  |
| **Make it Simple** |
| *Cuts through complexity to prioritise, clarify, and enable good decisions.* |
| **Lead Change Fearlessly** |
| *Drives transformation, embraces change, and helps others navigate uncertainty.* |
|  |
| **Learn and Adapt Fast** |
| *Experiments, reflects, and adjusts quickly to learn from experience.* |
|  |
| **Champion Diversity** |
| *Fosters inclusion, challenges bias, and ensures fairness in decision-making.* |

## 4. assessmentcenter\_data.csv

Contains the overall rating score for the sales assessment centers that are the fourth and final stage of the application process for entry level sales applicants.

|  |  |
| --- | --- |
| employee\_id | Unique identifier for every employee |
| job\_level | One of 5 job levels that define seniority in the organization (e.g. entry level) |
| department | The department an employee works in (e.g. Sales, HR etc.) |
| start\_month | The month of a particular year when an employee joined the company |
| status | The status assigned to every applicant at every assessment stage (hired, progressed or rejected) |
| stage | The application stage and the name of the application stage (e.g. stage 1: CV Screening comes before Stage 2: Telephone Interview |
| applicant\_id | Unique identifier for every applicant |
| job\_id | Unique identifier for every job that people applied and interviewed for |
| gender | Gender (Male, Female, Other) |
| assessmentcenter\_score | The overall score out of 5 they received in the assessment center. |

## 5. personality\_data.csv

Contains the scores on a personality assessment that applicants in entry-level positions go through as part of the assessment center and the outcome if they were hired or not. Unfortunately, there is no way to link this data to the assessment data due to data privacy guidelines (GDPR). Below is a brief description of the dimensions that are being assessed.

**Adjustment (inverse of Neuroticism):** Emotional stability, stress resilience, calmness. High scores = composed and calm under pressure.

**Ambition (Extraversion component):** Drive, goal orientation, leadership motivation. High scores = assertive, competitive.

**Sociability (Extraversion component):** Sociability, talkativeness, sociable nature. High scores = seeks social interaction, appears outgoing.

**Interpersonal Sensitivity (Agreeableness component):** Friendliness, tact, empathy. High scores = diplomatic, warm, considerate.

**Prudence (Conscientiousness component):** Conscientiousness, sense of duty, rule-following. High scores = organised, reliable, disciplined.

**Inquisitive (Openness component):** Intellectual curiosity, imagination, strategic thinking. High scores = imaginative, questions the status quo.

**Learning Approach (Openness component):** Learning orientation, interest in development and knowledge. High scores = eager to learn, committed to self-development.

## 5. employeesurvey\_data.csv

This dataset contains the scores for each question of the employee survey for every employee that took part in the survey last year (Beginning of April). Due to data protection guidelines it was not possible to get the employee ID since the survey team doesn’t want to erode trust in the anonymity. Therefore, they also ask you to not report any figures for groups smaller then 5 people! The survey system is however linked to the hr system so they could provide some of the variables you are already familiar with like employee status, performance etc. The dataset also includes the benchmark from the survey provider, since the items are standardized. People complain already about the survey lengths but the current items are also under review. You can suggest up to 3 additions max but ideally you also make suggestions for items to drop if you plan to add any items for the next round.