<https://rpubs.com/rhuebner/hrd_cb_v14>

Data Dictionary

| **Feature** | **Description** | **DataType** |
| --- | --- | --- |
| Employee Name | Employee’s full name | Text |
| EmpID | Employee ID is unique to each employee | Text |
| MarriedID | Is the person married (1 or 0 for yes or no) | Binary |
| MaritalStatusID | Marital status code that matches the text field MaritalDesc | Integer |
| EmpStatusID | Employment status code that matches text field EmploymentStatus | Integer |
| DeptID | Department ID code that matches the department the employee works in | Integer |
| PerfScoreID | Performance Score code that matches the employee’s most recent performance score | Integer |
| FromDiversityJobFairID | Was the employee sourced from the Diversity job fair? 1 or 0 for yes or no | Binary |
| Salary | The person’s yearly salary. $ U.S. Dollars | Float |
| Termd | Has this employee been terminated - 1 or 0 | Binary |
| PositionID | An integer indicating the person’s position | Integer |
| Position | The text name/title of the position the person has | Text |
| State | The state that the person lives in | Text |
| Zip | The zip code for the employee | Text |
| DOB | Date of Birth for the employee | Date |
| Sex | Sex - M or F | Text |
| MaritalDesc | The marital status of the person (divorced, single, widowed, separated, etc) | Text |
| CitizenDesc | Label for whether the person is a Citizen or Eligible NonCitizen | Text |
| HispanicLatino | Yes or No field for whether the employee is Hispanic/Latino | Text |
| RaceDesc | Description/text of the race the person identifies with | Text |
| DateofHire | Date the person was hired | Date |
| DateofTermination | Date the person was terminated, only populated if, in fact, Termd = 1 | Date |
| TermReason | A text reason / description for why the person was terminated | Text |
| EmploymentStatus | A description/category of the person’s employment status. Anyone currently working full time = Active | Text |
| Department | Name of the department that the person works in | Text |
| ManagerName | The name of the person’s immediate manager | Text |
| ManagerID | A unique identifier for each manager. | Integer |
| RecruitmentSource | The name of the recruitment source where the employee was recruited from | Text |
| PerformanceScore | Performance Score text/category (Fully Meets, Partially Meets, PIP, Exceeds) | Text |
| EngagementSurvey | Results from the last engagement survey, managed by our external partner | Float |
| EmpSatisfaction | A basic satisfaction score between 1 and 5, as reported on a recent employee satisfaction survey | Integer |
| SpecialProjectsCount | The number of special projects that the employee worked on during the last 6 months | Integer |
| LastPerformanceReviewDate | The most recent date of the person’s last performance review. | Date |
| DaysLateLast30 | The number of times that the employee was late to work during the last 30 days | Integer |
| Absences | The number of times the employee was absent from work. | Integer |

**Note that this dataset isn't perfect. By design, there are some issues that are present. It is primarily designed as a teaching data set - to teach human resources professionals how to work with data and analytics.**

### Content

We have reduced the complexity of the dataset down to a single data file (v14). The CSV revolves around a fictitious company and the core data set contains names, DOBs, age, gender, marital status, date of hire, reasons for termination, department, whether they are active or terminated, position title, pay rate, manager name, and performance score.

Recent additions to the data include:

* Absences
* Most Recent Performance Review Date
* Employee Engagement Score

### Inspiration

We've included some open-ended questions that you can explore and try to address through creating Tableau visualizations, or R or Python analyses. Good luck and enjoy the learning!

* Is there any relationship between who a person works for and their performance score?
* What is the overall diversity profile of the organization?
* What are our best recruiting sources if we want to ensure a diverse organization?
* Can we predict who is going to terminate and who isn't? What level of accuracy can we achieve on this?
* Are there areas of the company where pay is not equitable?
* What are age level categorization of employees hired this year?
* What management level get hired most this year.?
* Which month of year company has most joinings.?