Codebook for Attrition Dataset

# Data Overview

## Credentials

This data set can be downloaded from Kaggle data sets.

The direct link to data is: <https://www.kaggle.com/ludobenistant/hr-analytics/downloads/HR_comma_sep.csv>

## Business goal

This data was collected to answer the question:

Why are our best and most experienced employees leaving prematurely?

## Data description

This data set is a data frame of 10 variables over 15000 rows. Each row represents an employee that works in the company.

There are no missing values in the data.

# Variables description

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | Description | Type | Possible values |
| satisfaction\_level | Level of satisfaction (0-1) | Numeric | 0-1, 1 is high |
| last\_evaluation | Time since last performance evaluation (in Years) | Numeric | 0-1, 1 is high |
| number\_project | Number of projects completed while at work | Integer | 2, 3, 4, 5, 6, 7 |
| average\_montly\_hours | Average monthly hours at workplace | Integer | positive |
| time\_spend\_company | Number of years spent in the company | Integer | positive |
| Work\_accident | Whether the employee had a workplace accident | Integer | 1 = had an accident 0 = no accident |
| left | Whether the employee left the workplace or not (1 or 0) Factor | Integer | 1 = left 0 = works |
| promotion\_last\_5years | Whether the employee was promoted in the last five years | Integer | 1 = promoted 0 = not promoted |
| sales | Department in which they work for | Factor | 10 departments |
| salary | Relative level of salary (high) | Factor | High, low, medium |