**Comprender el dataset**

**Fuente de los datos**

La base de datos contiene registros de ausentismo en el trabajo del año 2024 de una empresa española de mensajería. Es una adaptación de la base de datos publicada en el enlace que está a continuación:

<https://archive.ics.uci.edu/dataset/445/absenteeism+at+work>

**Características y el alcance del dataset.**

* Es un conjunto de datos multivariado que cuenta con 845 registros y 21 columnas.
* Hay 31 duplicados iguales, que suponemos que deben ser un error de registro, y que, en consecuencia, vamos a eliminar.
* Cada registro representa la ausencia de un determinado empleado de la empresa en una fecha determinada.
* Tenemos 136 empleados que han registrado una ausencia en un año.

|  |  |  |  |
| --- | --- | --- | --- |
| Campo | Tipo | Descripción | Variación |
| ID | int | Número de Identificación del empleado | Estable |
| Reason\_absence | int | ICD (International Code Diseases): 28 posibles causas, ver Anexo. | Varía según la ausencia registrada |
| Month\_absence | int | Mes de la ausencia: del 1 al 12 | Varía según la ausencia registrada |
| Day\_week | int | Dia de la semana: Monday (2), Tuesday (3), Wednesday (4), Thursday (5), Friday (6) | Varía según la ausencia registrada |
| Seasons | int | Estación: summer (1), autumn (2), winter (3), spring (4) | Varía según la ausencia registrada |
| Transportation\_expense | int | Gastos de transporte en euros. No vemos ninguna correlación con la distancia ni con el número de ausencias, pero hay que seguir investigando | Estable |
| Distance\_Residence\_Work | int | Distancia residencia-trabajo (en Km) | Estable |
| Service\_time | int | Tiempo de servicio (en años). Creemos que es en años porque es estable y no varía según el mes que se registra la ausencia | Variación anual |
| Age | int | Edad del mensajero | Estable |
| Work\_load\_Average\_day | float | Carga de trabajo diario promedio (en minutos) | Varía según el día |
| Hit\_target | int | Objetivos alcanzados mensuales en porcentaje | Varía según el mes |
| Disciplinary\_failure | boolean | Faltas disciplinarias?: 0 (NO) ó 1 (SI) | Estable |
| Education | int | Educacion: high school (1), graduate (2), postgraduate (3), master and doctor (4) | Estable |
| Son | int | Número de hijos | Estable |
| Social\_drinker | boolean | Es un bebedor social?: 0 (NO) ó 1 (SI) | Estable |
| Social\_smoker | boolean | Es un fumador social?: 0 (NO) ó 1 (SI) | Estable |
| Pet | int | Número de mascotas | Estable |
| Weight | int | Peso (en Kg) | Estable |
| Height | int | Altura (en cm) | Estable |
| Body\_mass\_index | int | Indice de masa corporal | Estable |
| Absenteeism\_hours | int | Horas de ausentismo | Varía según la ausencia registrada |

**Características adicionales**

No hay valores nulos

**ANEXO: Causas de ausentismo**

Según el ICD (International Code Diseases) hay 26 posibles causas.

I Certain infectious and parasitic diseases

II Neoplasms

III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism

IV Endocrine, nutritional and metabolic diseases

V Mental and behavioural disorders

VI Diseases of the nervous system

VII Diseases of the eye and adnexa

VIII Diseases of the ear and mastoid process

IX Diseases of the circulatory system

X Diseases of the respiratory system

XI Diseases of the digestive system

XII Diseases of the skin and subcutaneous tissue

XIII Diseases of the musculoskeletal system and connective tissue

XIV Diseases of the genitourinary system

XV Pregnancy, childbirth and the puerperium

XVI Certain conditions originating in the perinatal period

XVII Congenital malformations, deformations and chromosomal abnormalities

XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified

XIX Injury, poisoning and certain other consequences of external causes

XX External causes of morbidity and mortality

XXI Factors influencing health status and contact with health services.

A esto, se agregan 7 categorías fuera del CID:  
  
patient follow-up (22),

medical consultation (23),

blood donation (24),

laboratory examination (25),

unjustified absence (26),

physiotherapy (27),

dental consultation (28).