

Absenteeism at work Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil.

Data Set Characteristics:	Multivariate, Time-Series	Number of Instances:	740	Area:	Business
Attribute Characteristics:	Integer, Real	Number of Attributes:	21	Date Donated	2018-04-05
Associated Tasks:	Classification, Clustering	Missing Values?	N/A	Number of Web Hits:	179659

Source:

Creators original owner and donors: Andrea Martiniano (1), Ricardo Pinto Ferreira (2), and Renato Jose Sassi (3).

E-mail address:

andrea.martiniano'@gmail.com (1) - PhD student;

log.kasparov'@gmail.com (2) - PhD student;

sassi'@uni9.pro.br (3) - Prof. Doctor.

Universidade Nove de Julho - Postgraduate Program in Informatics and Knowledge Management.

Address: Rua Vergueiro, 235/249 Liberdade, Sao Paulo, SP, Brazil. Zip code: 01504-001.

Website: <http://www.uninove.br/curso/informatica-e-gestao-do-conhecimento/>

Data Set Information:

The data set allows for several new combinations of attributes and attribute exclusions, or the modification of the attribute type (categorical, integer, or real) depending on the purpose of the research. The data set (Absenteeism at work - Part I) was used in academic research at the Universidade Nove de Julho - Postgraduate Program in Informatics and Knowledge Management.

Attribute Information:

1. Individual identification (ID)

2. Reason for absence (ICD).

Absences attested by the International Code of Diseases (ICD) stratified into 21 categories (I to XXI) as follows:

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.

And 7 categories without (CID) patient follow-up (22), medical consultation (23), blood donation (24), laboratory examination (25), unjustified absence (26), physiotherapy (27), dental consultation (28).

3. Month of absence
 4. Day of the week (Monday (2), Tuesday (3), Wednesday (4), Thursday (5), Friday (6))
 5. Seasons (summer (1), autumn (2), winter (3), spring (4))
 6. Transportation expense
 7. Distance from Residence to Work (kilometers)
 8. Service time
 9. Age
 10. Work load Average/day
 11. Hit target
 12. Disciplinary failure (yes=1; no=0)
 13. Education (high school (1), graduate (2), postgraduate (3), master and doctor (4))
 14. Son (number of children)
 15. Social drinker (yes=1; no=0)
 16. Social smoker (yes=1; no=0)
 17. Pet (number of pet)
 18. Weight
 19. Height
 20. Body mass index
 21. Absenteeism time in hours (target)

.arff header for Weka:

@relation Absenteeism_at_work

@attribute ID {31.0, 27.0, 19.0, 30.0, 7.0, 20.0, 24.0, 32.0, 3.0, 33.0, 26.0, 29.0, 18.0, 25.0, 17.0, 14.0, 16.0, 23.0, 2.0, 21.0, 36.0, 15.0, 22.0, 5.0, 12.0, 9.0, 6.0, 34.0, 10.0, 28.0, 13.0, 11.0, 1.0, 4.0, 8.0, 35.0}
 @attribute Reason_for_absence {17.0, 3.0, 15.0, 4.0, 21.0, 2.0, 9.0, 24.0, 18.0, 1.0, 12.0, 5.0, 16.0, 7.0, 27.0, 25.0, 8.0, 10.0, 26.0, 19.0, 28.0, 6.0, 23.0, 22.0, 13.0, 14.0, 11.0, 0.0}
 @attribute Month_of_absence REAL
 @attribute Day_of_the_week {5.0, 2.0, 3.0, 4.0, 6.0}
 @attribute Seasons {4.0, 1.0, 2.0, 3.0}
 @attribute Transportation_expense REAL
 @attribute Distance_from_Residence_to_Work REAL
 @attribute Service_time INTEGER
 @attribute Age INTEGER
 @attribute Work_load_Average/day_ REAL
 @attribute Hit_target REAL
 @attribute Disciplinary_failure {1.0, 0.0}
 @attribute Education REAL
 @attribute Son REAL
 @attribute Social_drinker {1.0, 0.0}
 @attribute Social_smoker {1.0, 0.0}
 @attribute Pet REAL
 @attribute Weight REAL
 @attribute Height REAL
 @attribute Body_mass_index REAL
 @attribute Absenteeism_time_in_hours REAL

Relevant Papers:

Martiniano, A., Ferreira, R. P., Sassi, R. J., & Affonso, C. (2012). Application of a neuro fuzzy network in prediction of absenteeism at work. In Information Systems and Technologies (CISTI), 7th Iberian Conference on (pp. 1-4). IEEE.

Citation Request:

Martiniano, A., Ferreira, R. P., Sassi, R. J., & Affonso, C. (2012). Application of a neuro fuzzy network in prediction of absenteeism at work. In Information Systems and Technologies (CISTI), 7th Iberian Conference on (pp. 1-4). IEEE.

Acknowledgements:

Professor Gary Johns for contributing to the selection of relevant research attributes.

Professor Emeritus of Management

Honorary Concordia University Research Chair in Management

John Molson School of Business

Concordia University

Montreal, Quebec, Canada

Adjunct Professor, OB/HR Division

Sauder School of Business,

University of British Columbia

Vancouver, British Columbia, Canada

Supported By:



In Collaboration With:



[About](#) || [Citation Policy](#) || [Donation Policy](#) || [Contact](#) || [CML](#)