Cascade Cup'20 Final Round

Problem Statement:

Absenteeism is a habitual pattern of absence from a duty or obligation without good reason. Generally, absenteeism is unplanned absences. If a workplace exhibits high degree of absenteeism there is a problem. It has been viewed as an indicator of poor individual performance, as well as a breach of an implicit contract between employee and employer.

This is a 740 x 22 tabular dataset, what each column represents is given in the data dictionary. Your task is to prepare an extensive report by preforming analysis on the dataset. Provide as many meaningful inferences and correlations using graphs as you can. Ensure that the report is not more than 10 pages (including the introduction and thank you pages).

Data Set Name:

Analysis of absenteeism in a company

Abstract:

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

Attribute Information:

1. Individual identification (ID)
2. Reason for absence (ICD).

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

> 1 Certain infectious and parasitic diseases
> 2 Neoplasms
> 3 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
> 4 Endocrine, nutritional and metabolic diseases
> 5 Mental and behavioral disorders
> 6 Diseases of the nervous system
> 7 Diseases of the eye and adnexa
> 8 Diseases of the ear and mastoid process
> 9 Diseases of the circulatory system
> 10 Diseases of the respiratory system
> 11 Diseases of the digestive system
> 12 Diseases of the skin and subcutaneous tissue
> 13 Diseases of the musculoskeletal system and connective tissue
> 14 Diseases of the genitourinary system
> 15 Pregnancy, childbirth and the puerperium
> 16 Certain conditions originating in the perinatal period
> 17 Congenital malformations, deformations and chromosomal abnormalities
> 18 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
> 19 Injury, poisoning and certain other consequences of external causes
> 20 External causes of morbidity and mortality
> 21 Factors influencing health status and contact with health services.

And 7 categories without (ICD), 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)