



دانشگاه صنعتی امیر کبیر
(پلی تکنیک تهران)

دانشکده ریاضی و علوم کامپیوتر

داده کاوی

دکتر فاطمه شاکری

پیش پردازش و تحلیل اکتشافی مجموعه داده

فروردین 1403

مجموعه داده ای که در اختیار شما قرار داده شده است ، اطلاعات جمع آوری شده از زمان های غیبت کارمندان شرکتی در بازه سال 2015 تا سال 2018 است.

هدف از تحلیل این مجموعه داده بررسی دلایل احتمالی غیبت کارمندان و برنامه ریزی های لازم جهت کاهش این ساعات است.

توضیحات مجموعه داده:

1. Absence_Reason (ICD)

غیبت های شهودی توسط کد بین المللی بیماری ها (کد بین المللی بیماری ها) به 21 دسته به شرح زیر تقسیم شده است:

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 behavioural 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

7 دسته بندی بدون کد ICD:

- 22. پیگیری بیمار
- 23. مشاوره پزشکی
- 24. اهدای خون
- 25. آزمایشگاه
- 26. غیبت غیر موجه
- 27. فیزیوتراپی
- 28. مشاوره دندان

2. Education_Level

- 1. دبیرستان
- 2. فارغ التحصیل
- 3. پسافارغ التحصیلی
- 4. ارشد و دکتری

3. Disciplinary_Incident , Social_Drinker_Status, Social_Drinker_Status

- 1: بله
- 0: خیر

نکات لازم:

- دادگان مورد نیاز در یک فایل اکسل در پیوست قرار گرفته است.
- فایل نوتبوک پروژه را میتوانید از لینک زیر دانلود کنید:
<https://colab.research.google.com/drive/130lhpJsJ45YeKzDfWddIJ60oyedjbmSR?usp=sharing>
- از شما انتظار میرود که به خواسته های مطرح شده در نوتبوک پاسخ دهید.
- در بخش EDA، تحلیل بصری لازم در بخش "توضیحات پاسخ" علاوه بر نمایش نمودار الزامی است.
- به صورت مناسب کامنت های لازم را در کدهای خود قرار دهید.
- فایل ارسالی خود را به صورت فشرده و به صورت "project_n_IDGroup.zip" نامگذاری کنید. فایل زیپ شامل یک فایل جویپتر نوت بوک و فایل حاوی لینک گوگل کولب باشد. دقت داشته باشید تنها فرستادن لینک گوگل کولب پذیرفته شده نیست. نیازی به ارسال گزارش به صورت مجزا نیست.
- آخرین مهلت ارسال تمرین، ساعت 23:59 روز سه شنبه 21 فروردین 1403 می باشد.