



دانشکده فنی و مهندسی
گروه مهندسی کامپیوتر

پروژه داده کاوی

نام دانشجو
صادق رجائی
شماره دانشجویی
40090992513

استاد درس
دکتر فاطمه باقری

نیمسال اول سال تحصیلی 1403-1404

1- مجموعه داده

مجموعه داده‌ها به موضوع تاثیر دورکاری بر سلامت روان می‌پردازد. این داده‌ها مربوط به پاسخ‌های افرادی است که در بازه‌های زمانی مختلف و در مکان‌های گوناگون به سوالاتی پیرامون سلامت روان و دورکاری پاسخ داده‌اند. این مجموعه شامل 5000 رکورد و 20 ویژگی است. ویژگی‌های این داده شامل عواملی مانند سن، جنسیت، وضعیت کاری، تأثیرات دورکاری بر سلامت روان، و میزان رضایت شغلی هستند. ویژگی‌ها دارای مقادیر null نیستند. در این مجموعه داده نقاط پرت نیز مشاهده می‌شود که ممکن است نیاز به توجه ویژه در فرآیند تحلیل داده‌ها داشته باشد.

1-1- مشخصات مجموعه داده

جدول توصیف مجموعه داده

ردیف	ویژگی	توصیف
1	Employee ID	شناسه منحصر به فرد هر کارمند
2	Age	سن هر کارمند
3	Gender	جنسیت هر کارمند
4	Job Role	موقعیت کاری هر کارمند
5	Industry	حوزه شغلی هر کارمند
6	Years of Experience	سابقه کار هر کارمند
7	Work Location	محل کار هر کارمند
8	Hours Worked Per Week	ساعات کاری هفتگی هر کارمند
9	Number of Virtual Meetings	تعداد ملاقات مجازی برقرار شده کارمند
10	Work Life Balance Rating	وضعیت تعادل بین زندگی شغلی و اجتماعی کارمند
11	Stress Level	میزان استرس هر کارمند
12	Mental Health Condition	وضعیت سلامت روان کارمند
13	Access to Mental Health Resources	آیا کارمند به امکانات بهبود سلامت روان دسترسی دارد؟
14	Productivity Change	تغییر تغییر کارآمدی هر کارمند
15	Social Isolation Rating	شدت انزوای محیط هر کارمند
16	Satisfaction with Remote Work	میزان رضایت از دور کاری برای هر کارمند
17	Company Support for Remote Work	میزان همکاری شرکت برای فراهم شدن دورکاری
18	Physical Activity	میزان فعالیت بدنی هر کارمند
19	Sleep Quality	وضعیت خواب هر کارمند
20	Region	منطقه جغرافیایی

1-2- نمونه هایی از مجموعه داده

Employee ID	Age	Gender	Job Role	Industry	Years of Experience
EMP0002	40	Female	Data Scientist	IT	3
EMP0135	60	Male	Marketing	Finance	2
EMP0137	44	Male	Data Scientist	Healthcare	14
EMP4891	50	Male	HR	Education	13

10	Finance	Designer	Female	47	EMP4887
7	Consulting	Project Manager	Non-binary	34	EMP4910

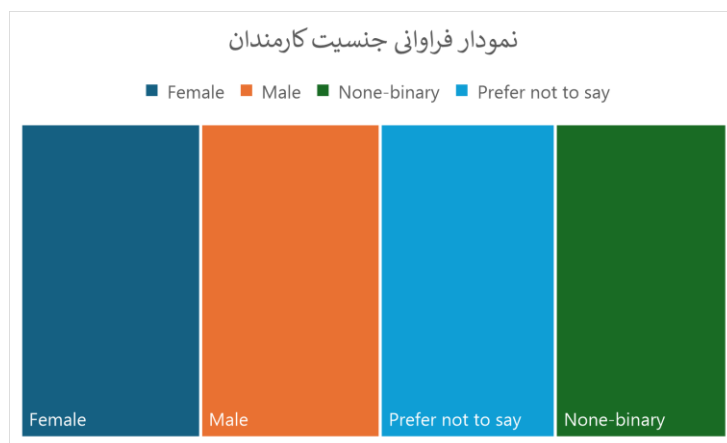
1-3- مشخصات اولیه ویژگی های مجموعه داده

ویژگی	تعداد	ماکزیم مقدار	مینیم مقدار	مقدار میانه	تعداد داده های ناموجود
Age	5000	60	22	41	0
Years of Experience	5000	35	1	18	0
Hours Worked Per Week	5000	60	20	40	0
Hours Worked Per Week	5000	15	0	8	0
Social Isolation Rating	5000	5	1	3	0
Company Support for Remote Work	5000	5	1	3	0
Gender	5000	Female → مورد 1274 Male → مورد 1270 Non-binary → مورد 1214 Prefer not to say → مورد 1242			0
Job Role	5000	Data Scientist → مورد 696 Designer → مورد 723 HR → مورد 716 Marketing → مورد 683 Project Manager → مورد 738 Sales → مورد 733 Software Engineer → مورد 711			0
Industry	5000	Consulting → مورد 680 Education → مورد 690 Finance → مورد 747 Healthcare → مورد 728 IT → مورد 746 Manufacturing → مورد 683 Retail → مورد 726			0
Work Location	5000	Hybrid → مورد 1649 Onsite → مورد 1637 Remote → مورد 1714			0
Stress Level	5000	High → مورد 1686 Low → مورد 1645 Medium → مورد 1669			0
Mental Health Condition	5000	Anxiety → مورد 1278 Burnout → مورد 1280 Depression → مورد 1246 None → مورد 1196			0
Satisfaction with Remote Work	5000	Neutral → مورد 1648 Satisfied → مورد 1677			0

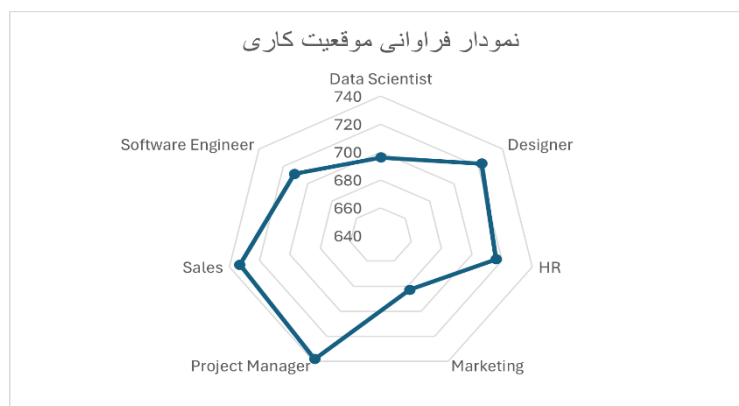
	Unsatisfied → مورد 1675		
0	Daily → مورد 1616 None → مورد 1629 Weekly → مورد 1755	5000	Physical Activity
0	Average → مورد 1628 Good → مورد 1687 Poor → مورد 1685	5000	Sleep Quality
0	Africa → مورد 860 Asia → مورد 829 Europe → مورد 840 North America → مورد 777 Oceania → مورد 867 South America → مورد 827	5000	Region

1-4- نمودارهای توصیف داده

در این بخش، نمودارهای توصیف داده‌ها جهت بررسی فراوانی و توزیع مقادیر ویژگی‌های مختلف ارائه شده‌اند. شکل ۱ فراوانی ویژگی جنسیت کارمندان را نشان می‌دهد، که شامل آمار متناسبی از جنسیت‌های متفاوت است. شکل ۲ به نحوه توزیع فراوانی موقعیت شغلی کارمندان اختصاص دارد که نمایانگر نسبت بیشتر مدیران پروژه می‌باشد. شکل ۳ و ۴ به توصیف فراوانی صناعی که کارمندان در آن مشغول به کار هستند می‌پردازد. شکل ۵ نمایانگر توزیع فراوانی سه شیوه کاری از حیث محل کار می‌باشد که مشخصاً با فاصله کمی تعداد موقعیت‌های دورکاری بیشتر می‌باشد. نمودارهای ۶، ۷ و ۸ به ترتیب توزیع گسسته سطح استرس، وضعیت روحی و روان و میزان رضایت از شیوه دورکاری پرداخته‌است.



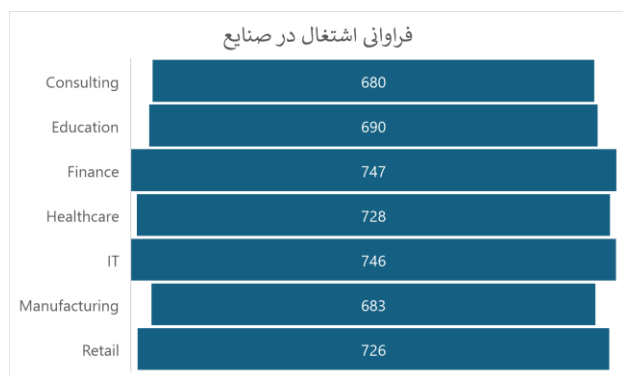
شکل 1: فراوانی ویژگی جنسیت در مجموعه داده



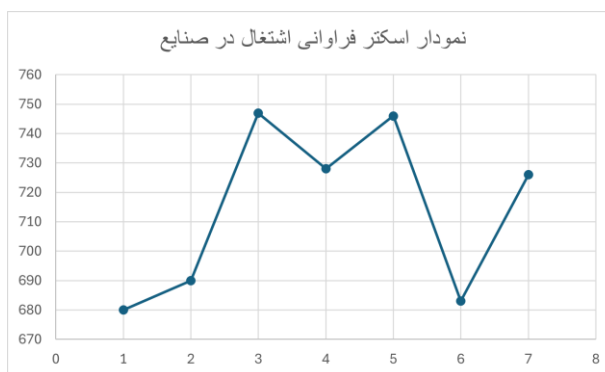
موقعیت کاری در

شکل 2: فراوانی

مجموعه داده

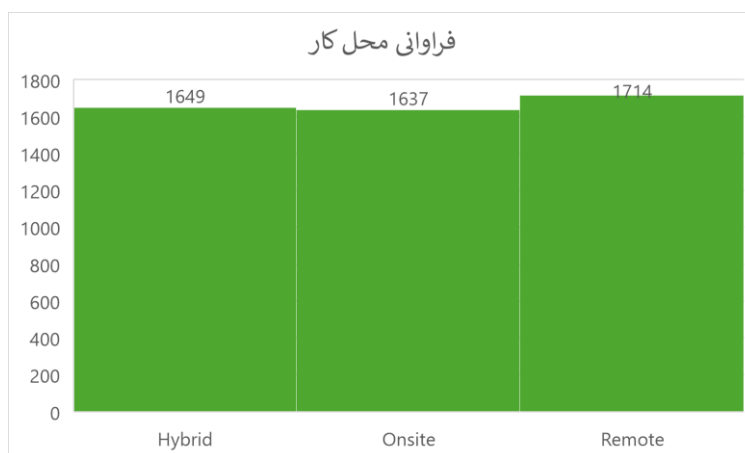


شکل 3: نمودار فراوانی اشتغال در

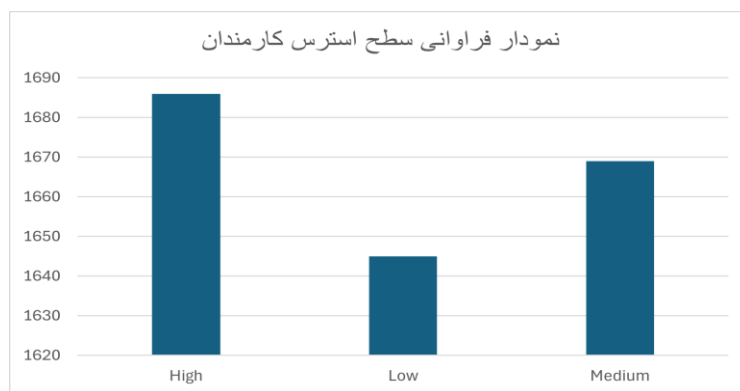


شکل 4: نمودار اسکتر اشتغال در صنایع

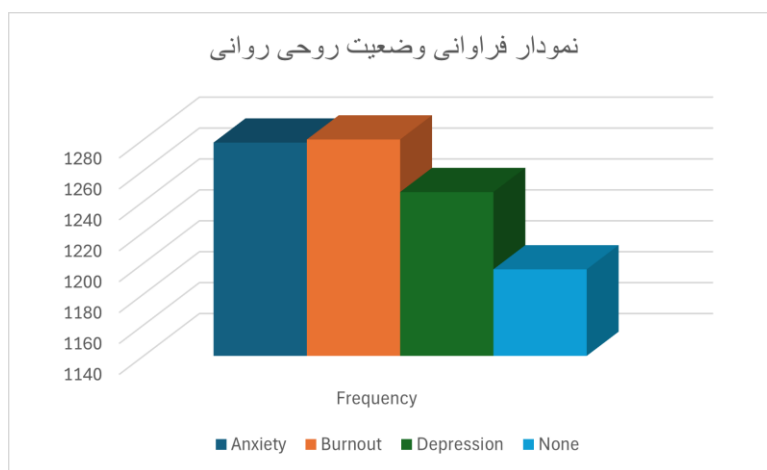
صنایع



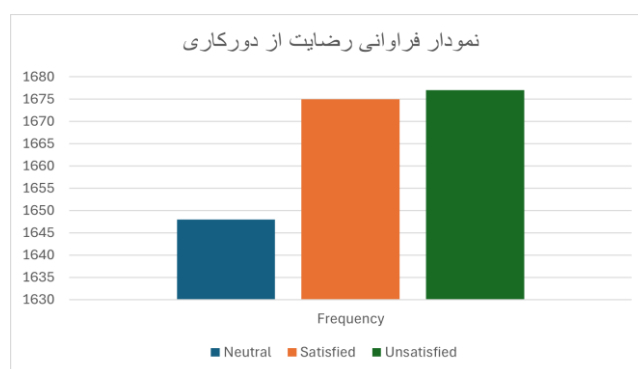
شکل 5: نمودار فراوانی محل کار



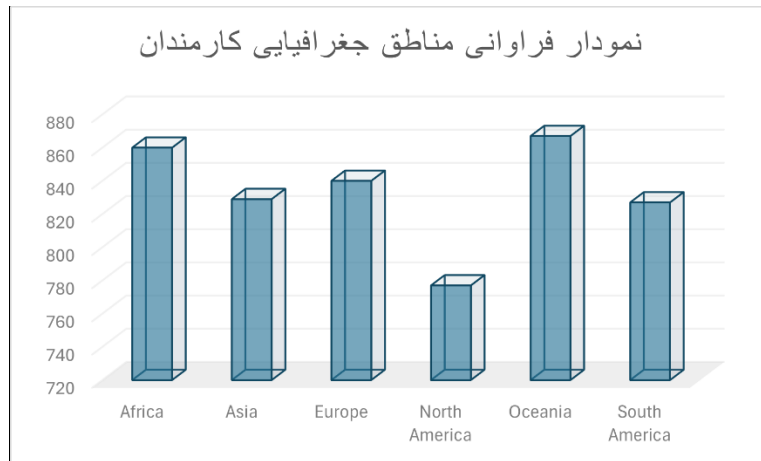
شکل 6: نمودار فراوانی سطح استرس کارمندان



شکل 7: نمودار فراوانی وضعیت روحی روانی



شکل 8: نمودار فراوانی رضایت از دورکاری



شکل 9: نمودار فراوانی مناطق جغرافیایی کارمندان

نقشه حرارتی

نقشه حرارتی زیر نشان‌دهنده همبستگی میان ویژگی‌های مختلف در مجموعه داده‌ای مرتبط با دورکاری و سلامت روان است. در این هیتمپ، هر خانه نمایانگر ضریب همبستگی بین دو ویژگی است که می‌تواند از -1 تا 1 متغیر باشد:

- همبستگی نزدیک به 1 نشان‌دهنده یک رابطه مثبت قوی است، به این معنا که افزایش یک ویژگی به طور قابل توجهی با افزایش ویژگی دیگر همراه است.
- همبستگی نزدیک به -1 نشان‌دهنده یک رابطه منفی قوی است، یعنی افزایش یک ویژگی با کاهش محسوس ویژگی دیگر همراه می‌شود.
- همبستگی نزدیک به 0 نشان‌دهنده نبود رابطه معنادار بین دو ویژگی است.



شکل 10: نقشه حرارتی برای همه ویژگی‌های اثر دورکاری بر سلامت روان