

学生生活中针对学业表现影响因素的数据统计分析与可视化

10225501437 寇璟奕


一、引言

现阶段，学生的学习压力越来越大，不少学生因压力过大抑或是成绩不理想出现抑郁、少言、不愿交友等现象。本实验旨在针对对于学业表现的影响因素，促进对特定背景下个人因素、学业成绩、心理健康、学习习惯和社会动态之间的相互作用进行全面分析。

二、数据来源与概述

来源：本数据集源于外网 [www.kaggle.com](https://www.kaggle.com/ismielhossenabir) 伊斯米尔·霍森·阿比尔博士的统计。

Collaborators

 Md. Ismiel Hossen Abir (Owner)

Authors

AUTHOR NAME	BIO
Md. Ismiel Hossen Abir	-

数据集（部分）如下：

	Age	Gender	AcademicPerformance	TakingNoteInClass	DepressionStatus	FaceChallengesToCompleteAcademicTask	LikePresentation	SleepPerDayHours	NumberOfFriend	LikeNewThings
1										
2	23	Male	Average	No	Sometimes	Yes	Yes	12		Yes
3	23	Male	Excellent	Sometimes	Yes	No	Yes	8	80	Yes
4	24	Male	Average	No	Sometimes	Sometimes	No	8	10	Yes
5	20	Female	Good	Yes	Sometimes	Yes	No	5	15	Yes
6	24	Female	Average	Yes	Yes	Yes	Yes	5	2	Yes
7	23	Male	Good	Sometimes	Sometimes	No	Yes	8	12	Yes
8	21	Female	Good	Yes	No	No	No	8	7	Yes
9	21	Male	Average	Sometimes	Yes	Sometimes	No	8	6	Yes
10	24	Male	Excellent	No	No	No	Yes	8	3	Yes
11	23	Female	Average	Yes	Sometimes	Sometimes	Yes	8	4	Yes
12	22	Male	Average	Yes	Yes	Sometimes	Yes	4	60	Yes
13	22	Male	Excellent	Sometimes	Sometimes	Yes	Yes	7	55	Yes
14	24	Male	Average	Yes	Sometimes	Sometimes	Yes	7	17	Yes
15	24	Female	Good	Yes	Yes	Yes	Yes	5	3	Yes

该数据集随机调查了 100 位计算机专业学生的年龄、性别、学业成绩、笔记习惯、抑郁状态、是否遇到挑战、喜好演示（衡量外向或是内向）、睡眠时间、朋友数量、新事物接受度这几个因素。

在本实验中，我们重点讨论年龄、性别、学业成绩、笔记习惯、抑郁程度、睡眠时间、朋友数量这几个因素间是否相互影响及其相互关系。

三、数据分析与可视化

1、单变量分析

我先以学业情况作为应变量,其他因素作为自变量,探查其他因素对于学业情况的影响,这里主要以生成箱形图/条形图/散点图为主，便于我们看清两个变量间的关系。

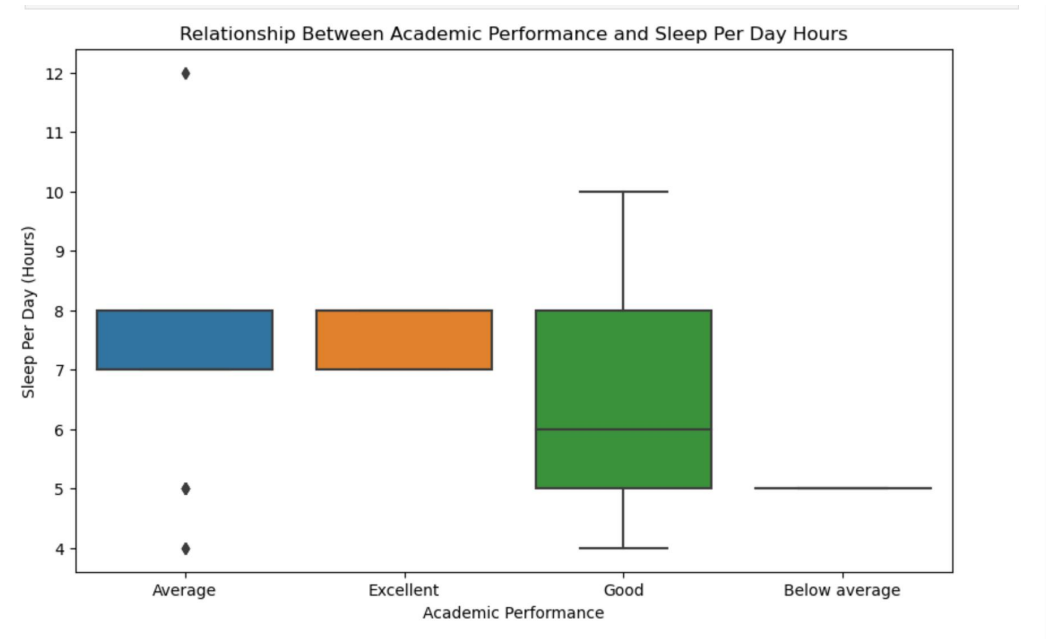
1、睡眠时长与学业情况

根据箱形图我们可以看到：学业表现“优秀”的学生群体普遍显示出较高的睡眠时长，且睡眠时长分布较为集中。这表明在这个群体中，充足的睡眠可能与优秀的学业表现相关联；

学业表现为“良好”的学生的睡眠时长略低于“优秀”群体，但仍然保持在较高水平，表明良好的睡眠习惯可能也有助于良好的学业成绩。

对于学业表现“平均”的学生，睡眠时长的分布较广，中位数处于中等水平。这意味着在这个群体中，睡眠时长对学业表现的影响不如前两组明显。

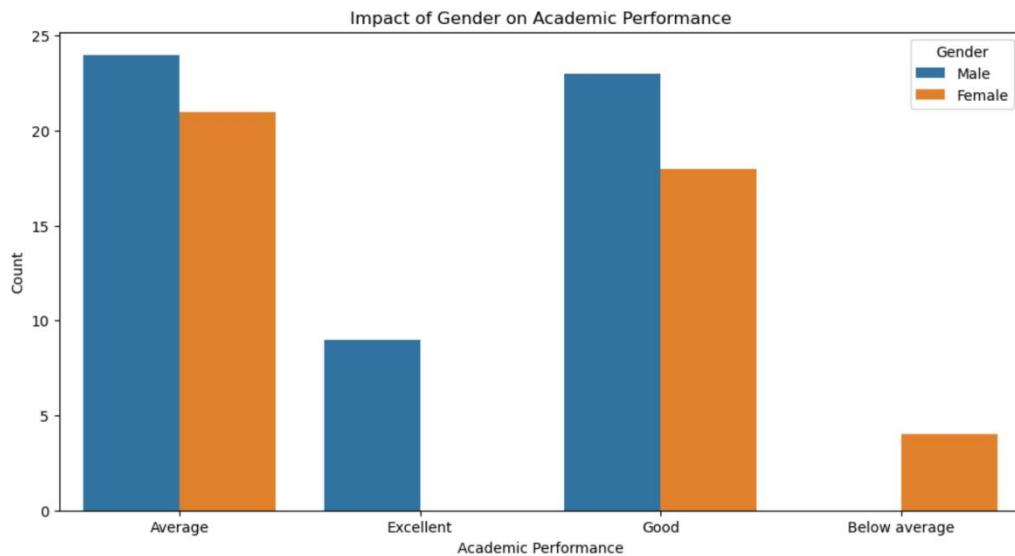
学业表现“低于平均”的学生群体中，睡眠时长普遍较短，我认为这暗示着睡眠不足与较差的学业表现之间可能存在某种关联。



综上所述我们可以得到：睡眠时长与学业表现有一定的相关性，其中睡眠充足的学生倾向于有更好的学业表现。

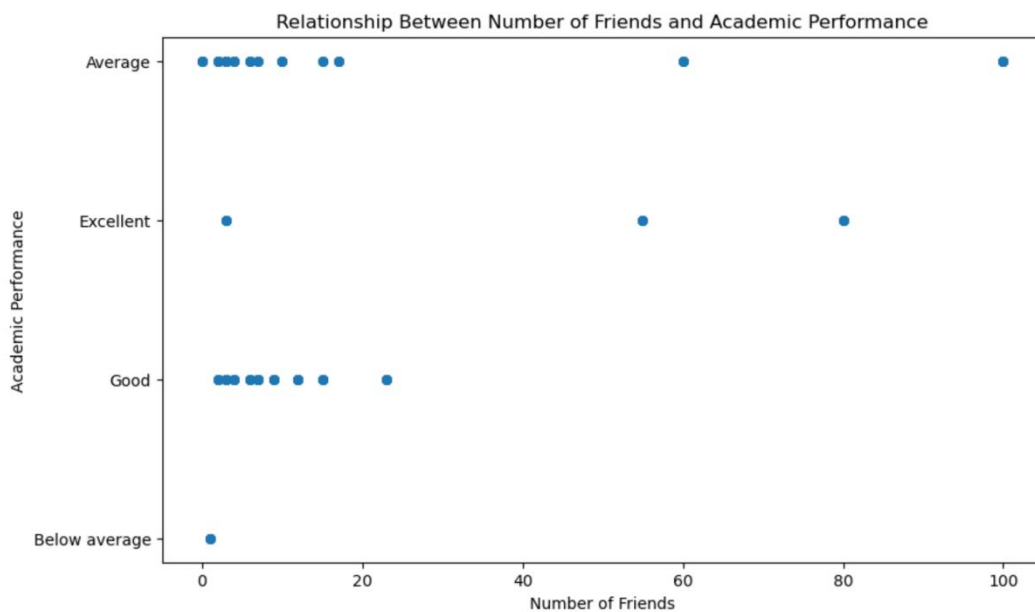
2、性别与学业情况

基于条形图：女性学生在“优秀”类别中的比例略高于男性，说明在这一学业表现最高层级中，女性学生的表现略优于男性；男性学生在“良好”类别中稍微多于女性学生，说明在这一层级，男性和女性的学业表现相对接近，但男性略占优势；在“平均”这个类别中，男性和女性学生数量相近，说明在平均水平的学业表现中，两性之间没有显著差异；而在“低于平均”学业表现中：男性学生数量显著多于女性，说明在低于平均水平的学业表现中，男性学生的比例较高。



综上所述：性别与学习情况有一定关联。女性在最高学业表现等级中略占优势，而男性在低于平均等级的学业表现中比例较高。

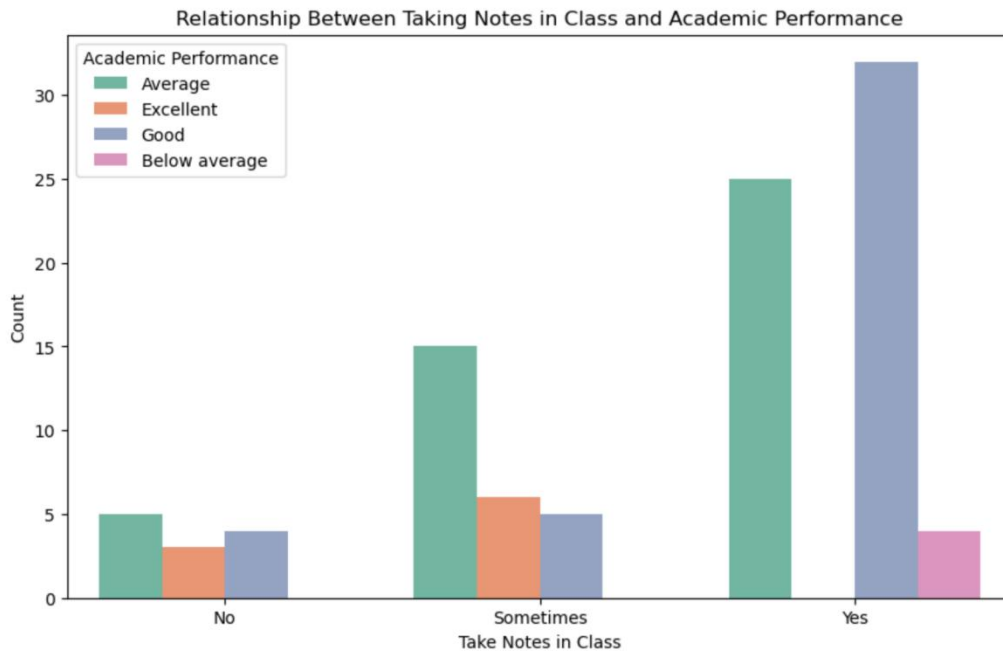
3、朋友数量与学业表现



由散点图可见：对于“优秀”、“良好”、“平均”等学业评级，朋友数量均分布在一个较宽的范围内变化，整体上，二者没有明确关联或趋势。说明朋友数量与学业表现没有直接关联性。

4、笔记习惯与学业情况

基于条形图可知：对于"Excellent"和"Good"学业表现的学生，更多的学生倾向于在课堂上做笔记，而"Average"学业表现的学生中，做笔记和不做笔记的学生数量相对均衡。



这表明，在某些情况下，做笔记与较好的学业表现之间存在正相关关系。

11、多变量分析

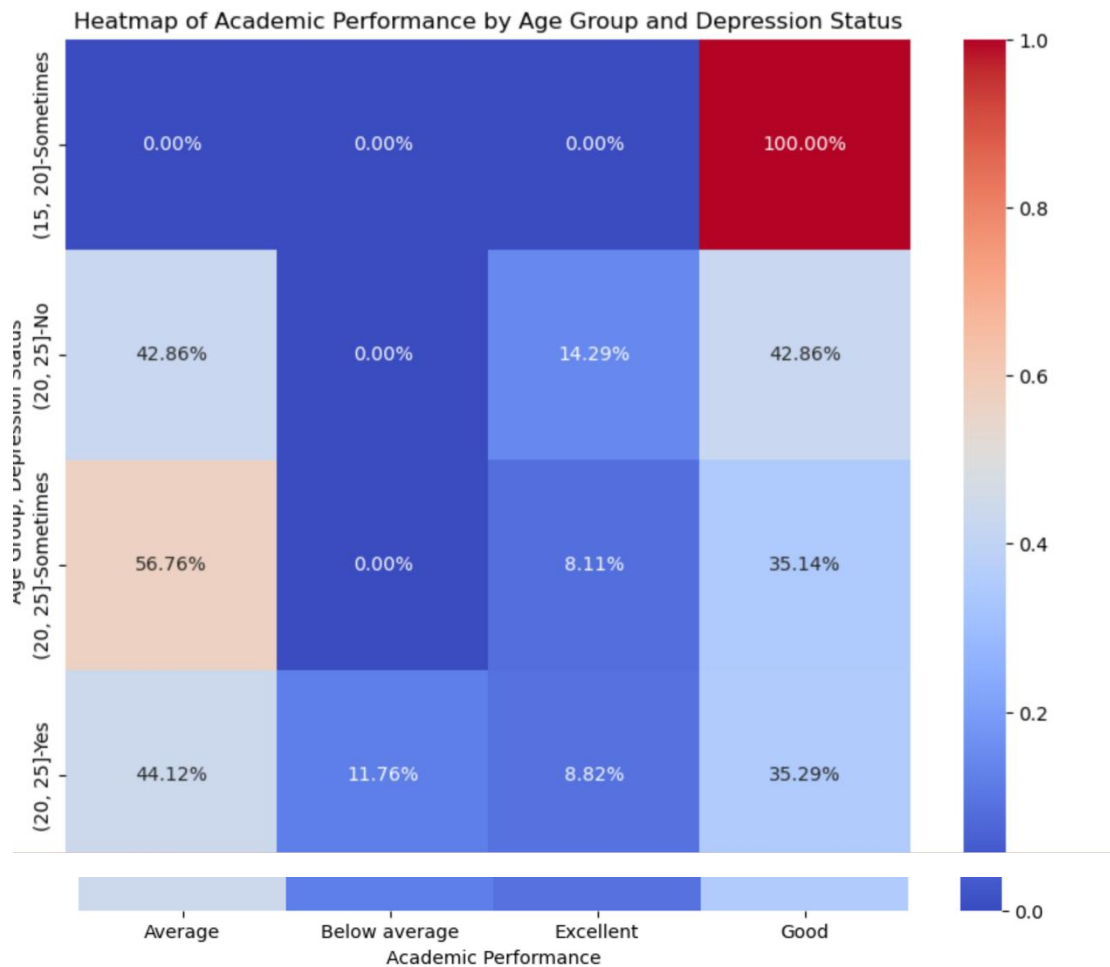
进一步考察，探究多变量间相互的关系。这里不再明确规定自变量与应变量。

1、基于热力图讨论年纪、抑郁状态与学习情况间关系

由热力图可见：在年龄较低的组别(例如 15-20 岁)，不同抑郁状态下的学生在学业表现上的分布比较均匀。在这个年龄组中，抑郁状态对学业表现的影响似乎不是特别显著。

随着年龄的增长，特别是在 20-25 岁和 25-30 岁的年龄组，抑郁状态“有时”和“总是”的学生在学业表现“平均”和“低于平均”的比例有所增加。这表明在这些年龄组中，抑郁状态可能对学业表现有更明显的影响。

在 30-35 岁和 35-40 岁的年龄组中，抑郁状态“有时”和“总是”的学生在较低学业表现等级上的比例相对较高，尤其是在“平均”和“低于平均”的类别中。



热力图显示了随着年龄的增长，抑郁状态对学业表现的影响变得更加显著，尤其是在表现较差的学生群体中。

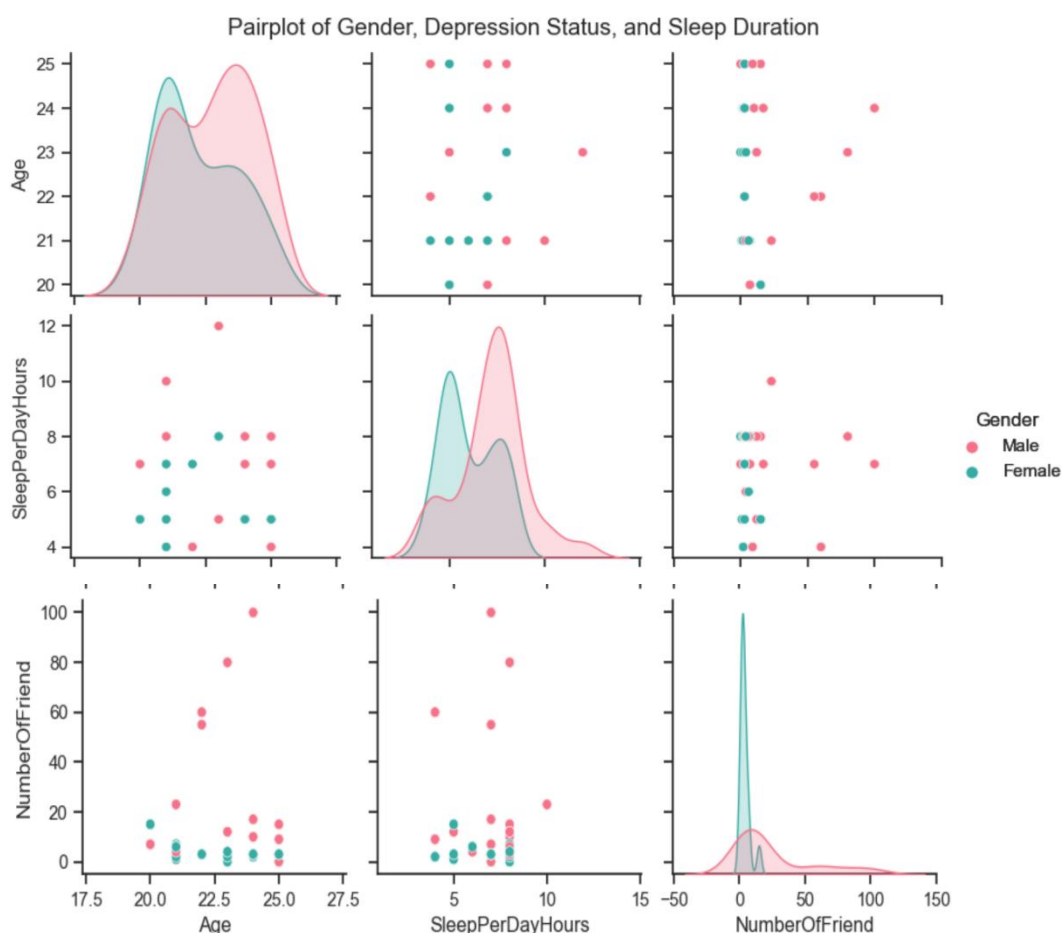
2、基于成对图讨论性别、抑郁状态、睡眠时长关系

从成对图中可以观察到，性别与抑郁状态之间没有明显的直接关系，分布状态相似。性别与睡眠时长之间分布均匀，也没有明显的直接关系。

抑郁状态与睡眠时长间：低抑郁状态低抑郁状态 (Low Depression Status) 态的学生通常呈现出较长的睡眠时长分布。这意味着低抑郁状态的学生可能更倾向于拥有较长的每日睡眠时间。

中等抑郁状态 (Moderate Depression Status)：中等抑郁状态的学生在睡眠时长上呈现出中等范围的分布。他们的睡眠时长可能分布在中间范围，没有明显的偏向长或短。

高抑郁状态 (High Depression Status)：高抑郁状态的学生在睡眠时长上通常呈现出较短的睡眠时长分布。这意味着高抑郁状态的学生可能更倾向于拥有较短的每日睡眠时间。



综上，抑郁状态与睡眠时长之间的关联趋势是：随着抑郁状态从低到高的增加，学生的平均睡眠时长可能呈现出逐渐减少的趋势。这意味着在样本中，高抑郁状态的学生可能更容易出现较短的睡眠时长，而低抑郁状态的学生可能更容易拥有较长的睡眠时长。

四、讨论与总结

通过数据分析，我们了解到睡眠时长、性别、笔记习惯、抑郁状态都会对学业水平造成影响，此影响随学习水平或是年纪的不同有高低，且睡眠时长与抑郁状态二者也彼此相关。

为了让学生们有一个健康的校园生活和社会心理，我们要注重对于学生们心理状态的辅导，及时疏解压力消极情绪，减少抑郁的发生，从而让学生们有一个相对良好的睡眠，培养出健康的社会心理，这样才会更加有动力在学业上拼搏。学生们也可以培养自己记笔记的习惯，作为提升自己学习效率的一种方式。同时，我们也注意到，在本实验中，朋友数量与学习情况、抑郁状态并不直接挂钩，所以一些同学们不必为显得合群而刻意去迎合别人交友，保持自己的本心行动即可。

希望此实验能够对了解学生学习生活与社会心理有所帮助。