

Analyse und Visualisierung von medizinischen Daten im Bereich der psychischen Gesundheit mit Python

Efim Shliamin

July 7, 2023

1 Einleitung

In den letzten Jahren hat das Bewusstsein für psychische Gesundheit und deren Auswirkungen auf verschiedene Lebensbereiche zugenommen. Besonders im Bildungsbereich spielt die psychische Gesundheit von Studierenden eine bedeutende Rolle. Beispielsweise heißt es in Artikel "Grading bias and young adult mental health", dass sich Bewertungsverzerrungen auf die Gesundheit und die Kompetenzentwicklung in den Jahren nach dem Abschluss auswirken, die Auswirkungen jedoch je nach Geschlecht, sozioökonomischem Hintergrund und Migrationshintergrund unterschiedlich sind, was bedeutet, dass Bewertungsverzerrungen sich auf gesundheitliche Ungleichheiten auswirken können [1].

Darüber hinaus heißt es in Artikel "Effects of Mental Health on Student Learning", dass es festgestellt wurde, dass psychische Erkrankungen mit einem geringeren akademischen Erfolg und einem schlechteren Studienabschluss zusammenhängen [2].

Eine gute psychische Gesundheit kann sich positiv auf den akademischen Erfolg auswirken, während psychische Belastungen und Probleme negative Auswirkungen haben können.

Der vorliegende Datensatz, "A Statistical Research on the Effects of Mental Health on Students' CGPA", wurde durch eine Umfrage mittels Google Forms unter Studierenden einer Universität erhoben. Das Hauptziel dieser Studie bestand darin, den Zusammenhang zwischen der mentalen Gesundheit von Studierenden und ihrem akademischen Erfolg, gemessen an ihrem kumulativen Notendurchschnitt (CGPA), zu untersuchen. Wir haben die Ergebnisse unserer Studie mit den oben zitierten Ergebnissen verglichen und festgestellt, dass in unserer Studie kein negativer Zusammenhang zwischen Noten und psychischen Störungen bei Schülern besteht, unsere Studie jedoch auf der Unter-

suchung der psychischen Gesundheit von Schülern vor ihrem Abschluss basiert Universitäten.

Die Umfrage umfasste Fragen zu verschiedenen Aspekten der psychischen Gesundheit, wie Stress, Angstzustände, Depressionen und Schlafstörungen. Darüber hinaus wurden demografische Informationen und Daten zum akademischen Hintergrund der Studierenden erfasst.

Durch die Analyse dieses Datensatzes können wichtige Erkenntnisse darüber gewonnen werden, wie sich die psychische Gesundheit auf den akademischen Erfolg von Studierenden auswirkt. Diese Frage ist umfassend und hilft uns, die Ergebnisse unserer Studie mit der in der Zeitschrift Health Economics veröffentlichten Studie zu vergleichen. Die Ergebnisse könnten sowohl für Bildungseinrichtungen als auch für Psychologen und Berater von Interesse sein, um gezielte Unterstützungsmaßnahmen und Interventionen zu entwickeln, die die psychische Gesundheit und das Wohlbefinden der Studierenden fördern.

Im Rahmen dieser Projektarbeit werden wir den Datensatz analysieren und mithilfe statistischer Methoden untersuchen, wie sich verschiedene Faktoren der psychischen Gesundheit auf den CGPA der Studierenden auswirken. Darüber hinaus werden wir die Ergebnisse visualisieren, um die Zusammenhänge besser darzustellen und zu interpretieren.

Das Projekt hat das Potenzial, einen Beitrag zur Verbesserung des Verständnisses der psychischen Gesundheit von Studierenden zu leisten und Empfehlungen für Bildungseinrichtungen zur Förderung des Wohlbefindens und des akademischen Erfolgs von Studierenden abzuleiten.

2 Datensammlung

Dieser Fragebogen wurde auf der Grundlage des Datensatzes "A statistical research on the effects of mental health on students' CGPA dataset" entwickelt und dient der Erfassung von Informationen über die Auswirkungen der psychischen Gesundheit auf den akademischen Erfolg von Studenten.

2.1 Fragebogen

Frage	Ja	Nein
Haben Sie eine Diagnose einer Depression?		
Haben Sie Symptome von Angstzuständen?		
Haben Sie Panikattacken?		
Haben Sie einen Spezialisten für eine Behandlung aufgesucht?		

Dieser Datensatz wurde von 100 Studenten gesammelt durch eine von Google Forms durchgeführte Umfrage unter Universitätsstudenten, um ihre aktuelle akademische Situation und psychische Gesundheit zu untersuchen.

3 Vorläufige Datenanalyse

Der Datensatz enthält die folgenden Merkmale:

- Zeitstempel (*Timestamp*): Ein Zeitstempel, der das Datum und die Uhrzeit der Umfrageerfassung angibt. Das Format ist "Tag/Monat/Jahr Stunde:Minute:Sekunde".
- Geschlecht (*Choose your gender*): Eine diskrete Variable, die das Geschlecht der Person angibt. Die Werte können "Männlich" oder "Weiblich" sein.
- Alter (*Age*): Eine kontinuierliche Variable, die das Alter der Person in Jahren angibt.
- Kurs/Spezialisierung (*What is your course?*): Eine diskrete Variable, die den aktuellen Kurs oder die Spezialisierung der Person beschreibt.
- Aktuelles Studienjahr (*Your current year of Study*): Eine diskrete Variable, die das aktuelle Studienjahr der Person angibt.
- Durchschnittsnote (*What is your CGPA?*): Eine diskrete Variable, die den Durchschnittsnotenwert der Person angibt. Die Antworten werden in Form von Notenbereichen angegeben, z.B. "3,00 - 3,49".
- Familienstand (*Marital status*): Eine diskrete Variable, die den Familienstand der Person angibt.
- Depression (*Do you have Depression?*): Eine diskrete Variable, die angibt, ob die Person an Depressionen leidet. Die Antworten können "Ja" oder "Nein" sein.
- Angst (*Do you have Anxiety?*): Eine diskrete Variable, die angibt, ob die Person unter Angstzuständen leidet. Die Antworten können "Ja" oder "Nein" sein.
- Panikattacken (*Do you have Panic attack?*): Eine diskrete Variable, die angibt, ob die Person an Panikattacken leidet. Die Antworten können "Ja" oder "Nein" sein.
- Behandlung durch Spezialisten (*Did you seek any specialist for a treatment?*): Eine diskrete Variable, die angibt, ob die Person eine Behandlung durch einen Facharzt gesucht hat. Die Antworten können "Ja" oder "Nein" sein.

Um mit Daten in Python zu arbeiten, einschließlich des Lesens und Filterns von Daten sowie mathematischen Operationen für die Arbeit mit Datenarrays, haben wir die Frameworks und Bibliotheken Pandas und NumPy verwendet. Wir haben auch die Bibliothek Matplotlib verwendet, um die erhaltenen Ergebnisse in Python zu visualisieren. Nachdem die Datenaufbereitung abgeschlossen war, haben wir die folgenden Schritte durchgeführt:

1. Filtering: die Daten wurden analysiert und keine anomalen Werte wurden festgestellt, die die Analyseergebnisse beeinflussen könnten. Diese Überprüfung

war erforderlich, um genauere Ergebnisse zu gewährleisten.

2. Behandlung fehlender Werte: Es wurden keine fehlenden Werte in den Daten gefunden, die die Genauigkeit der Analyse beeinträchtigen könnten. Daher musste man diese Lücken nicht durch Durchschnittswerte oder am häufigsten vorkommende Werte in den entsprechenden Datenfeldern auffüllen.

3. Kodierung kategorischer Variablen: Der Datensatz enthielt kategorische Variablen wie Geschlecht, Bildungsrichtung und Familienstand. Um sie in der Analyse verwenden zu können, hat man sie mit der Methode des One-Hot-Encodings in numerisches Format umgewandelt.

4. Entfernen überflüssiger Variablen: Nach der Datenanalyse wurden die Merkmale ignoriert, die für diese Studie nicht relevant waren, wie zum Beispiel der Familienstand und die Antwort auf die Frage, ob der Student oder die Studentin einen Spezialisten zur Behandlung sucht.

Diese Schritte der Datenvorbereitung haben uns geholfen, die Daten für weitere Analysen vorzubereiten und zuverlässigere Ergebnisse in dieser Studie über medizinische Daten zum psychischen Wohlbefinden zu erzielen.

4 Datenanalyse und Datenvisualisierung

Zunächst war es wichtig herauszufinden, in welchem Alter Depression bei Studenten am häufigsten auftritt. Dazu wurde eine Analyse der Häufigkeit von Depressionen in Bezug auf die von den Studenten angegebenen Alterswerte durchgeführt. Die Ergebnisse der Daten sind in Diagramm 1 dargestellt.

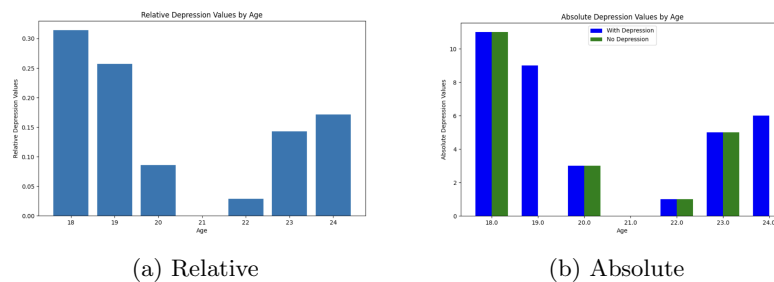


Figure 1: Relative (a) and Absolute (b) Depression Values by Age

Aus den Diagrammen geht hervor, dass mehr als 30 Prozenten aller depressiven Erkrankungen bei Studierenden im Alter von 18 Jahren auftreten.

Als nächster Schritt war es interessant, die Abhängigkeit von Angststörungen von den Altersgruppen der Studenten zu untersuchen. Die Ergebnisse der Daten sind in Diagramm 2 dargestellt.

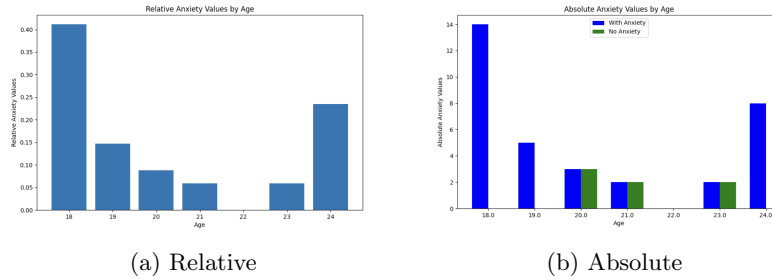


Figure 2: Relative (a) and Absolute (b) Anxiety Values by Age

Aus den Diagrammen geht hervor, dass mehr als 40 Prozenten aller depressiven Erkrankungen bei Studierenden im Alter von 18 Jahren auftreten.

Als dritten Schritt war es wichtig herauszufinden, in welchem Alter Panikattacken bei Studenten am häufigsten auftreten. Die Ergebnisse der Daten sind in Diagramm 3 dargestellt.

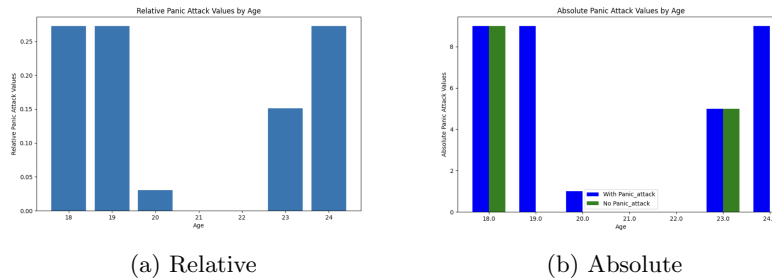


Figure 3: Relative (a) and Absolute (b) Panic Attack Values by Age

Aus den Diagrammen geht hervor, dass mehr als ein Viertel aller Fälle von Panikattacken bei Studierenden im Alter von 18, 19 und 24 Jahren auftreten.

Anschließend haben wir eine Korrelationsanalyse durchgeführt, um festzustellen, ob es einen Zusammenhang zwischen dem Alter der Studenten und ihrem Studienjahr gibt. Die Ergebnisse sind in Abbildung 4 dargestellt.

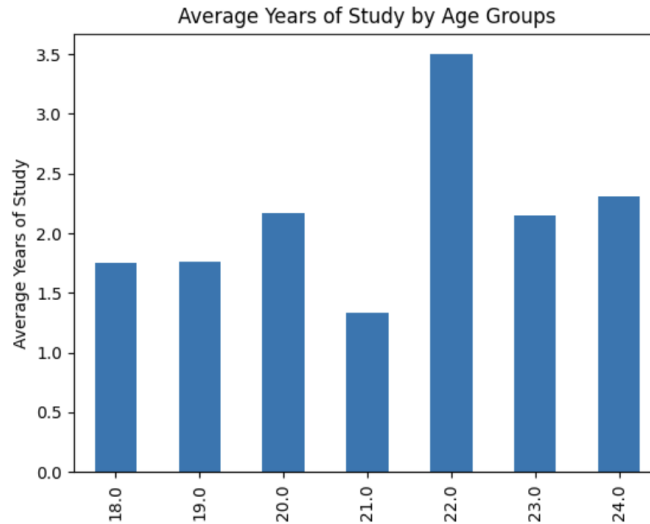


Figure 4: Zusammenhang zwischen Alter und Studienjahr

Aus dem Diagramm lässt sich ableiten, dass wir einen allgemeinen Trend gefunden haben: Im Alter von 22 Jahren befindet sich die Mehrheit der Studierenden bereits am Ende ihres Studiums. Dafür haben wir die Methode zur Erstellung eines Säulendiagramms (Bar Plot) verwendet.

Basierend auf den durchgeführten Analysen wurde eine Untersuchung zur Leistung der Studierenden in Abhängigkeit von psychischen Störungen durchgeführt. Dabei wurden statistische Verfahren wie der t-Test angewendet (siehe Punkt 4.1).

Die Daten wurden in zwei Gruppen aufgeteilt: Studierenden mit psychischen Störungen und Studierenden ohne psychische Störungen. Die Nullhypothese lautete, dass es keinen Unterschied in der Leistung zwischen den beiden Gruppen gibt, während die Alternativhypothese davon ausging, dass ein Unterschied besteht.

Es ist jedoch wichtig zu beachten, dass andere Einflussfaktoren wie Bildungsniveau, sozioökonomischer Status oder das Vorliegen anderer gesundheitlicher Probleme ebenfalls berücksichtigt werden sollten.

4.1 Zusammenhänge finden

Das Ziel unserer Analyse ist es, Zusammenhänge zwischen den Noten der Studenten und ihren psychischen Störungen zu identifizieren. Wir verwenden statistische Methoden, um Muster und Beziehungen zwischen den Noten und den psychischen Zuständen der Studenten zu erkennen. Durch die Untersuchung dieser

Zusammenhänge können wir mögliche Auswirkungen der psychischen Gesundheit auf die akademische Leistung der Studierenden verstehen.

Dabei wurden statistische Verfahren wie der t-Test angewendet. Die Ergebnisse liefern folgende Informationen:

Durchschnittsnote für Studierende mit Depression: 3.142857142857143 Dieser Wert repräsentiert den Durchschnitt der Noten (CGPA) für Studierende, die angegeben haben, an Depressionen zu leiden.

Durchschnittsnote für Studierende ohne Depression: 3.0454545454545454 Dieser Wert repräsentiert den Durchschnitt der Noten (CGPA) für Studierende, die keine Depressionen angegeben haben.

Statistischer Test: $t = 0.6544866337067948$, $p = 0.514315972959327$ Der t-Wert ist ein Maß für den Unterschied der Durchschnittswerte zwischen den beiden Gruppen (Studierende mit Depression und Studierende ohne Depression). In diesem Fall beträgt der t-Wert 0.6544866337067948. Der p-Wert beträgt 0.514315972959327 und gibt die Wahrscheinlichkeit an, die Daten zu beobachten, wenn es keinen wirklichen Unterschied zwischen den beiden Gruppen gibt.

5 Interpretation der Ergebnisse

Basierend auf den Durchschnittsnoten haben Studierende mit Depressionen einen etwas höheren Durchschnitt (3.142857142857143) im Vergleich zu Studierenden ohne Depressionen (3.0454545454545454). Die Ergebnisse des t-Tests mit einem t-Wert von 0.6544866337067948 und einem p-Wert von 0.514315972959327 deuten jedoch darauf hin, dass es keinen signifikanten Unterschied in den Durchschnittsnoten zwischen den beiden Gruppen gibt. Der p-Wert von 0.514315972959327 ist größer als das üblicherweise verwendete Signifikanzniveau von 0.05, was darauf hindeutet, dass der beobachtete Unterschied wahrscheinlich auf den Zufall zurückzuführen ist und nicht auf einen tatsächlichen Unterschied.

6 Schlussfolgerung

P-Wert (p-value) ist ein statistisches Maß, das uns hilft, die Evidenz gegen die Nullhypothese zu bewerten. Im Rahmen eines statistischen Tests gibt das p-Wert an, wie wahrscheinlich es ist, die beobachteten Daten oder noch extremere Ergebnisse zu erhalten, wenn die Nullhypothese wahr ist. Ein kleiner p-Wert deutet darauf hin, dass die beobachteten Daten unwahrscheinlich sind, wenn die Nullhypothese wahr ist, und unterstützt die Alternative.

T-Wert (t-value) ist ein Maß für statistische Signifikanz, das in Hypothesentests verwendet wird. Er zeigt, wie stark sich die Durchschnittswerte zweier Gruppen (z. B. Durchschnittsnoten von Studenten mit und ohne Depression)

in Standardfehlern unterscheiden. Ein höherer t-Wert deutet auf eine größere Abweichung zwischen den Gruppen hin.

Für diese Arbeit haben wir die Durchschnittsnoten von Studenten mit und ohne Depression verglichen. Die Ergebnisse zeigen, dass die Durchschnittsnote der Studenten mit Depression (3.142857142857143) etwas höher ist als die Durchschnittsnote der Studenten ohne Depression (3.0454545454545454). Der t-Wert beträgt 0.6544866337067948 und das p-Wert beträgt 0.514315972959327. Da das p-Wert größer als das übliche Signifikanzniveau (z. B. 0.05) ist, haben wir keine ausreichenden Beweise, um die Nullhypothese abzulehnen. Das bedeutet, dass wir keinen signifikanten Unterschied in den Durchschnittsnoten zwischen den beiden Gruppen feststellen konnten. Es kann also nicht eindeutig gesagt werden, dass die Depression einen signifikanten Einfluss auf die Noten hat.

Daher widersprechen unsere Ergebnisse der Aussage, die in dem Artikel "Effects of Mental Health on Student Learning" veröffentlicht wurde.

References

- [1] Ulf-G. Gerdtham Gawain Heckley Anna Linder, Martin Nordin. Grading bias and young adult mental health. *Health Econ.*, 2022.
- [2] Ren VanderLind. Effects of mental health on students learning.