

Bachelorarbeit  
im Bachelorstudiengang  
**Betriebswirtschaft**  
an der Hochschule für angewandte Wissenschaften Neu-Ulm

**Untersuchung von Täuschungsverhalten in digitalen Prüfungen  
Identifikation von Einflussfaktoren mithilfe von maschinellem Lernen**

Erstkorrektorin: Prof. Dr. Sonja Köppl  
Zweitkorrektor: Prof. Dr. Marlon Füller

Verfasserin: Stefanie Wittgens (Matrikel-Nr.: 273130)

Thema erhalten: 27.10.2022  
Arbeit abgegeben: 27.02.2023

## Inhaltsverzeichnis

Abbildungsverzeichnis .....	III
Tabellenverzeichnis .....	V
Abkürzungsverzeichnis .....	VI
1. Einleitung .....	1
1.1. Das Projekt ii.oo.....	1
1.2. Zielsetzung .....	2
1.3. Aufbau der Arbeit.....	2
2. Maschinelles Lernen .....	3
2.1. Definition und Abgrenzung.....	3
2.2. Arten des maschinellen Lernens .....	5
2.2.1. Überwachtes Lernen.....	6
2.2.2. Unüberwachtes Lernen.....	12
3. Datenerhebung .....	14
4. Auswertung mit Python .....	16
4.1. Aufbereitung der Daten.....	16
4.2. Algorithmen und Methoden in Python zur Auswertung.....	18
4.2.1. LabelEncoder und OneHotEncoder .....	19
4.2.2. StandardScaler .....	20
4.2.3. PCA.....	20
4.2.4. KMeans .....	21
5. Interpretation der Ergebnisse .....	21
5.1. Deskriptive Auswertung .....	21
5.2. Clusterauswertung .....	32
6. Fazit und Empfehlung .....	40
Literaturverzeichnis.....	VII
Anhang .....	IX

## Abbildungsverzeichnis

Abbildung 2-1 Abgrenzung der Begriffe (In Anlehnung an: Wuttke o. J., Abschnitt: Was ist Deep Learning?).....	5
Abbildung 2-2 Übersicht des Maschinellen Lernens und seine Untergruppen (In Anlehnung an: Marco Arnold o.J.).....	6
Abbildung 2-3 Darstellung lineare Trennung (Hirschle 2021, S. 63).....	8
Abbildung 2-4 Beispiel der Nearest Neighbour-Methode (Ertel 2021, S. 218) ....	8
Abbildung 2-5 Beispiel k-Nearest Neighbour-Methode für k-nächste Nachbarn (Ertel 2021, S. 220).....	9
Abbildung 2-6 Entscheidungsbaum (In Anlehnung an Papp et al. 2022, 229) ..	11
Abbildung 2-7 Beispiel für Support Vector Machine mit Klassengrenze (Papp et al. 2022, S. 227) .....	11
Abbildung 2-8 Darstellung eines Dendrogramms (Ertel 2021, S. 264) .....	13
Abbildung 4-1 Ausgabe der "describe()" -Funktion aus Python (Eigene Berechnung) .....	18
Abbildung 4-2 OneHotEncoding der Spalten mit string Werten (Eigene Darstellung) .....	19
Abbildung 4-3 Normalisierung des DataFrames (Eigene Darstellung).....	20
Abbildung 4-4 Code der PCA (Eigene Darstellung).....	20
Abbildung 4-5 Code des kMeans Algorithmus (Eigene Darstellung) .....	21
Abbildung 5-1 Verteilung der Studiengänge (Eigene Darstellung).....	22
Abbildung 5-2 Geschlechtsverteilung der Umfrageteilnehmer (Eigene Darstellung) .....	23
Abbildung 5-3 Wohnsituation der Umfrageteilnehmer (Eigene Darstellung) .....	23
Abbildung 5-4 Auswertung der Zufriedenheit mit dem Studium und der Hochschule mittels Likert-Skala (Eigene Darstellung).....	24
Abbildung 5-5 Auswertung des Leistungsdrucks und der Motivation der Befragten mittels Likert-Skala (Eigene Darstellung) .....	25
Abbildung 5-6 Freizeitbeschäftigung der Befragten (Eigene Darstellung).....	26
Abbildung 5-7 Regelmäßiger Besuch der Vorlesung (Eigene Darstellung).....	26
Abbildung 5-8 Angabe, bei welcher Prüfung sich die Befragten wohler fühlen (Eigene Darstellung) .....	27
Abbildung 5-9 Angabe, welche Prüfung Befragte als fairer empfinden (Eigene Darstellung) .....	27

Abbildung 5-10 Angaben zur Teilnahme an online Prüfungsformen.....	28
Abbildung 5-11 Angabe bei welchen Online Prüfungen geschummelt wurde (Eigene Darstellung) .....	29
Abbildung 5-12 Haltung zum Schummeln bei Online Prüfungen (Eigene Darstellung) .....	29
Abbildung 5-13 Antworten zur Frage, welche Konsequenzen Befragte vom Schummeln abhalten (Eigene Darstellung) .....	30
Abbildung 5-14 Verwendung des Spickzettels in verschiedenen Prüfungen (Eigene Darstellung) .....	31
Abbildung 5-15 Verwendung der Kommunikation mit anderen während der Prüfung als Täuschungsmittel in verschiedenen Prüfungen (Eigene Darstellung) .....	31
Abbildung 5-16 kMeans Clustering mit drei Clustern (Eigene Darstellung).....	32
Abbildung 5-17 Studiengang-Auswertung der Cluster (Eigene Darstellung).....	33
Abbildung 5-18 Geschlechtsverteilung der Cluster (Eigene Darstellung) .....	34
Abbildung 5-19 Wohnorte in den Clustern (Eigene Darstellung) .....	35
Abbildung 5-20 Clusterauswertung der Zufriedenheit mit dem Studium (Eigene Darstellung) .....	36
Abbildung 5-21 Clusterauswertung der Zufriedenheit mit der Hochschule (Eigene Darstellung) .....	37
Abbildung 5-22 Clusterauswertung der Motivation (Eigene Darstellung).....	37
Abbildung 5-23 Clusterauswertung des Leistungsdrucks (Eigene Darstellung) 38	
Abbildung 5-24 Clusterauswertung der Freizeitaktivitäten (Eigene Darstellung) .....	39
Abbildung 5-25 Clusterauswertung der Schummel Haltung und des tatsächlichen Schummel Verhalten (Eigene Darstellung) .....	39

## Tabellenverzeichnis

Tabelle 4-1 Aufschlüsselung der Ersetzungen der Zeitangaben .....	17
Tabelle 4-2 Aufschlüsselung der Notenersetzung .....	18
Tabelle 5-1 Tabelle der Studiengänge mit Fakultäten (Eigene Darstellung) .....	22
Tabelle 5-2 Bekanntheit der Schummel Methoden (Eigene Darstellung) .....	30
Tabelle 5-3 Mittelwert und Standardabweichung des Notendurchschnitts, des Fachsemesters und des Alters aller Daten und der Clusterzugehörigen Teilnehmer (Eigene Darstellung) .....	34
Tabelle 6-1 Zusammenfassung der Clusterergebnisse (Eigene Darstellung) ...	41

## Abkürzungsverzeichnis

BGW	Betriebswirtschaft im Gesundheitswesen
BWL	Betriebswirtschaft
DEM	Digital Enterprise Management
DMP	Digitales Medizin- und Pflegemanagement
DP	Digitale Prüfung
DSM	Data Science Management
GM	Fakultät Gesundheitsmanagement
GPM	Game-Produktion und Management
ICA	independent component analysis (Unabhängigkeitsanalyse)
IG	Informationsmanagement im Gesundheitswesen
IM	Fakultät Informationsmanagement
IMA	Information Management Automotive
IMUK	Informationsmanagement und Unternehmenskommunikation
KNN	künstliche neuronale Netze
k-NN	k-Nearest-Neighbor
M	Mittelwert
m	Männlich
max	Maximum
min	Minimum
PCA	principal component analysis (Hauptkomponentenanalyse)
PHA	Physician Assistant
SD	Standardabweichung
SVM	Support Vector Machine
w	weiblich
WIF	Wirtschaftsinformatik
WIN	Wirtschaftsingenieurwesen
WSP	Wirtschaftspsychologie
WW	Fakultät Wirtschaftswissenschaften
Zoom	Zoom Video Communications (Software Unternehmen für Videokonferenzen)

## 1. Einleitung

Die Corona Pandemie im Jahr 2020 forderte von vielen Hochschulen und Universitäten eine schnelle Umstrukturierung ihrer Lehre und ihres Prüfungswesens. Um dieser gerecht zu werden, wurden Lehr- und Prüfveranstaltungen online abgehalten.<sup>1</sup> Dieser Umstand der Veränderung stellt nicht nur die Hochschulen und Universitäten vor Herausforderung. Studierende und Lehrende selbst hatten enorme Herausforderungen zu stemmen. Die online Lehre zwingt Studierende zu einem isolierten Lernen und einer höheren psychischen Belastung.<sup>2</sup>

Diese Herausforderungen beeinflussten Studierende in ihrem Verhalten und ihrer Einstellung gegenüber akademischer Integrität. Eine Studie aus dem Jahr 2021 befasst sich mit unehrlichem Verhalten während des Sommersemesters 2020 und stellte fest, dass 61,4% der Befragten bei Online Prüfungen geschummelt haben. Auch stellte die Studie fest, dass drei der fünf meisten Antworten beim Schummel Verhalten direkt mit Online Prüfungen in Zusammenhang stehen.

Um das Schummeln in Online Prüfungen zu verhindern, gibt es mehrere potenzielle Möglichkeiten in dieser Studie. Die Anpassung der Prüfungen, das stärkere Überwachen der Prüfungen und die Belohnung von tieferem Verständnis der Lerninhalte.<sup>3</sup> All diese Möglichkeiten hängen mit der Veränderung der Prüfung und Prüfungssituation zusammen.

Um diese Möglichkeiten weiter zu untersuchen, wurde das Projekt ii.oo ins Leben gerufen.

### 1.1. Das Projekt ii.oo

Das Projekt „ii.oo digitales kompetenzorientiertes prüfen“ wird im Rahmen der Ausschreibung „Innovationen in der Hochschullehre“ gefördert und von neun Hochschulen in Bayern durchgeführt. Die teilnehmenden Hochschulen sind Coburg, Regensburg, Landshut, München, Augsburg, Neu-Ulm, Nürnberg und Würzburg-Schweinfurt. Die Idee des Projektes ist, die Lehre so zu gestalten das Prüfungssysteme chancengleich, rechtlich und technisch sicher sind. Außerdem

---

<sup>1</sup> vgl. Goertz und Hense 2021, S. 6.

<sup>2</sup> vgl. Breitenbach 2021, 8f.

<sup>3</sup> vgl. Janke et al. 2021.

soll die Haltung der Studierenden und Lehrenden gefördert werden, damit ein selbstverantwortliches Lernen möglich ist. Die Kompetenzorientierung soll das Leitbild der Lehre werden. Verschiedene Teilziele des Projektes sind die Entwicklung neuer, weiterentwickelter Prüfungsformen, die Erstellung von allgemeingültigen Leitlinien für digitale Prüfungen, die Untersuchung der Haltung der Studierenden und Lehrenden zum selbstverantwortlichen Lernen und Verminderung von Selbst- und Fremdtäuschungen und die Weiterentwicklung von Prüfungssysteme mit sicheren und skalierbaren Lösungen für digitale kompetenzorientierte Prüfungen.<sup>4</sup>

Im Rahmen der Untersuchung zur Haltung der Studierenden und Lehrenden zum selbstverantwortlichen Lernen und der Verminderung von Selbst- und Fremdtäuschungen wurde das Thema dieser Bachelorarbeit erarbeitet.

## 1.2. Zielsetzung

Das Ziel der Arbeit ist verschiedene Einflussfaktoren von Studierenden aus ihrem Privatleben und dem schulischen Umfeld auf ihre Leistungen zu identifizieren. Außerdem soll ausgewertet werden, ob diese Einflussfaktoren einen Zusammenhang mit unehrlichem Verhalten während Online Prüfungen aufweisen. Die Auswertung soll mithilfe von maschinellem Lernen unterstützt werden. Dabei soll herausgefunden werden, ob es Gruppen von Einflussfaktoren gibt, die zu einem erhöhten Schummel Verhalten bei Online Prüfungen führen. Durch diese Analyse sollen Studierende besser verstanden werden und Handlungsempfehlungen entstehen, die zu einem besseren selbstverantwortlichem Lernen und der Vermeidung von Selbst- und Fremdtäuschen führen können.

## 1.3. Aufbau der Arbeit

In dieser Arbeit werden verschiedene Einflussfaktoren der Studierenden untersucht, um Verhaltensweisen aufzuzeigen, die eine höhere Wahrscheinlichkeit von unehrlichem Verhalten in Online Prüfungen aufweisen. Nachfolgend wird die Vorgehensweise in der Arbeit genauer erläutert.

---

<sup>4</sup> vgl. Kormann et al. 2022.

Das erste Kapitel dieser Arbeit umfasst eine Einleitung, die Zielsetzung und die Vorgehensweise. Des Weiteren wird das Projekt ii.oo vorgestellt.

Im zweiten Kapitel der Arbeit liegt der Fokus auf dem maschinellen Lernen. Es wird dargestellt aus welchen Unterbegriffen es besteht und wie diese Unterbegriffe in verschiedene Methoden der Auswertung aufgeteilt sind. Zudem wird erklärt, wie diese Methoden funktionieren und für welche Daten sie am besten geeignet sind. Dies dient zum besseren Verständnis der Arbeit.

Im weiteren Teil der Arbeit wird das methodische Vorgehen bei der Umfrage-Erstellung und -Auswertung vorgestellt. Es folgt eine Erklärung der wichtigsten Codes in Python<sup>5</sup> zum Verständnis der Auswertung. Danach werden die Ergebnisse der Umfrage dargestellt und im weiteren Verlauf mithilfe von Python und maschinellen Lernen ausgewertet, veranschaulicht und in Beziehung gesetzt.

Abschließend gibt es eine Zusammenfassung der wichtigsten Erkenntnisse mit entsprechenden Handlungsempfehlungen und ein Ausblick für das weitere Vorgehen der Untersuchung.

## 2. Maschinelles Lernen

Mit diesem Kapitel wird das Ziel verfolgt, eine Einführung in das maschinelle Lernen zu geben. Zuerst wird auf die Definition des maschinellen Lernens und die Abgrenzung zu anderen verwendeten Begriffen eingegangen. Danach werden verschiedene Arten des maschinellen Lernens erläutert und dargestellt für welche Datenmenge sie geeignet sind beziehungsweise welches Ergebnis mit ihnen erzielt werden können. Zuletzt wird dargestellt, wie Daten vorbereitet werden können.

### 2.1. Definition und Abgrenzung

Maschinelles Lernen wird in der Literatur häufig auch mit dem englischen Begriff Machine Learning beschrieben. Dabei handelt es sich um einen Teilbereich der

---

<sup>5</sup> Python ist eine universelle und gut lesbare Programmiersprache.

künstlichen Intelligenz und Robotik, der sich mit dem Lernen aus Erfahrung beschäftigt. Vorbild hierfür ist das menschliche und tierische Lernen. Die theoretische Grundlage von maschinellem Lernen ist die Statistik, da aus Stichproben und Daten Schlussfolgerungen gezogen werden. Dabei erkennen sie Regel- und Gesetzmäßigkeiten in den verwendeten Daten und können diese nach ausreichendem Training auf andere Daten übertragen und auswerten. Beispiele von bereits funktionierenden Algorithmen des maschinellen Lernens sind die Gesichts- und Spracherkennung. Die bekanntesten Beispiele für die Spracherkennung sind Siri von Apple und Alexa von Amazon. Diese Systeme können gesprochene Wörter umwandeln um maschinell Befehle auszuführen. Die erkannten Muster und Regelmäßigkeiten helfen den Prozess zu verstehen um Vorhersagen in der nahen Zukunft treffen zu können. Ein Modell das Vorhersagen trifft wird prädiktiv genannt. Wohingegen ein Modell das Wissen aus Daten generiert, deskriptiv genannt wird. Ein Modell kann auch beide Formen nutzen.<sup>6 7</sup>

Das maschinelle Lernen wird in verschiedene Bereiche unterteilt. Die Bereiche sind: das unüberwachte Lernen, das überwachte Lernen und das verstärkte Lernen. Die Bereiche werden im nachfolgende Unterkapitel weiter erläutert.<sup>8</sup>

An dieser Stelle muss angemerkt werden, dass bei der Recherche von maschinellem Lernen die Begriffe künstliche Intelligenz, Deep Learning und künstliche neuronale Netze (KNN) zu finden sind. Die Abgrenzung dieser Begriffe müssen gesondert differenziert betrachtet werden.

Bei der Künstlichen Intelligenz werden Regeln festgelegt und von Algorithmen abgearbeitet, wohingegen beim maschinellen Lernen die Regeln mithilfe der Beispiele gelernt werden. Deshalb ist die künstliche Intelligenz ein Überbegriff von maschinellem Lernen (siehe Abbildung 2-1).<sup>9</sup>

---

<sup>6</sup> vgl. Bendel 2019.

<sup>7</sup> vgl. Alpaydın 2022, 1ff.

<sup>8</sup> vgl. Alpaydın 2022, S. 4–16.

<sup>9</sup> vgl. Papp et al. 2022, S. 220.

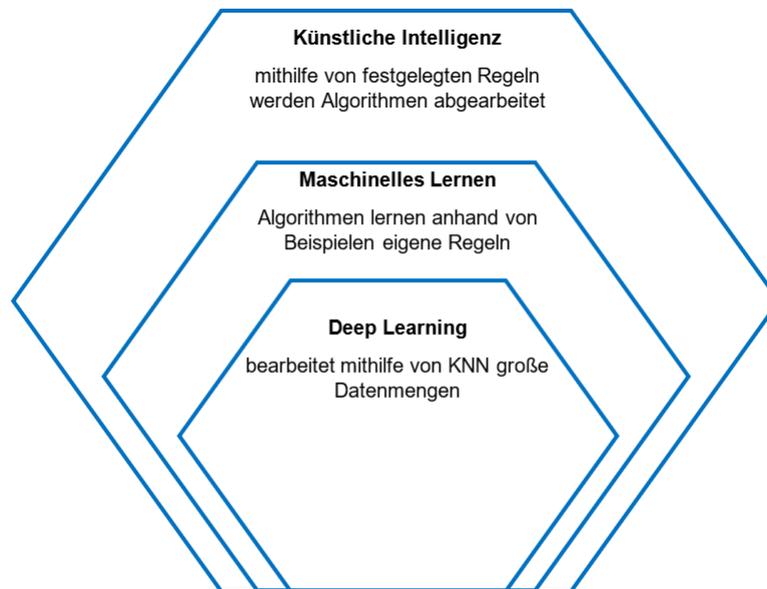


Abbildung 2-1 Abgrenzung der Begriffe (In Anlehnung an: Wuttke o. J., Abschnitt: Was ist Deep Learning?)

Das Deep Learning stellt ein Teilgebiet des maschinellen Lernens da. Dabei verwendet das Deep Learning eine sehr große Datenmenge und greift zur Bearbeitung ausschließlich auf künstliche neuronale Netze zurück.<sup>10</sup>

## 2.2. Arten des maschinellen Lernens

Nachdem im letzten Kapitel auf die Definition des maschinellen Lernens und die Abgrenzung zu anderen Begriffen eingegangen wurde, werden in diesem Kapitel die verschiedenen Arten des maschinellen Lernens beleuchtet. Das maschinelle Lernen kann dabei in zwei große Unterbegriffe eingeteilt werden (siehe Abbildung 2-2): das überwachte Lernen (Supervised Learning) und das unüberwachte Lernen (Unsupervised Learning). Diese Untergruppen beinhalten verschiedene Arten von Algorithmen, die in den nachfolgenden Unterkapitel weiter erläutert werden.<sup>11</sup>

<sup>10</sup> vgl. Wuttke o. J., Absatz: Was ist der Unterschied zwischen Deep Learning und Maschine Learning?

<sup>11</sup> vgl. Nguyen und Zeigermann 2021, S. 3–5.

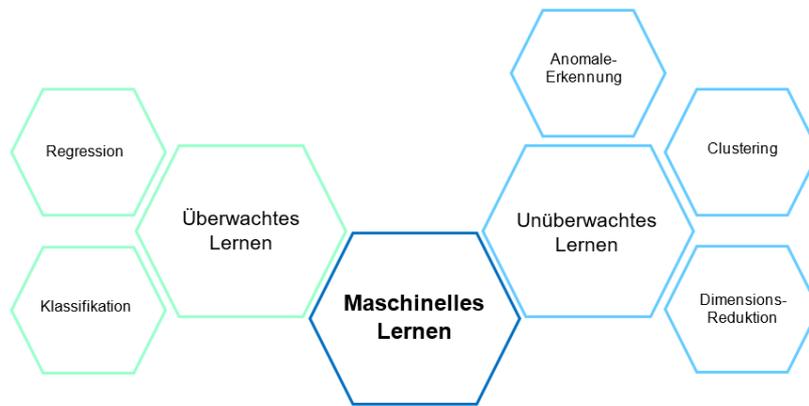


Abbildung 2-2 Übersicht des Maschinellen Lernens und seine Untergruppen (In Anlehnung an: Marco Arnold o.J.)

Beim überwachtem Lernen greift der Algorithmus auf sogenannte Labels (Kennungen) zurück. Mithilfe der Labels kann der Algorithmus seine Performance bei einer Aufgabe verbessern. Der Algorithmus ist aufgrund der Labels in der Lage die Daten einzuordnen. Beispielsweise beim Erkennen von Spam Emails, anhand der Labels können die Emails bereits in Spammails und normalen Emails unterteilt werden. Der Algorithmus des unüberwachten Lernens hat keinen Zugriff auf Labels. Er löst Aufgaben indem er ähnliche Daten in eine Gruppe ordnet, diese Gruppe sich aber von anderen Gruppen unterscheidet. Unüberwachte Probleme sind weniger klar definiert und deshalb schwieriger zu lösen, jedoch ist die Lösung am Ende leistungsfähiger. Aufgrund der fehlenden Aufgabe für den Algorithmus, kann dieser mehr Muster finden als gesucht. Deshalb sind unüberwachte Systeme flexibler und besser beim Finden von neuen Mustern in zukünftigen Daten.<sup>12</sup>

## 2.2.1. Überwachtes Lernen

### 2.2.1.1. Regression

Die Regression modelliert Beziehungen zwischen verschiedenen Variablen. Dies kann durch unterschiedliche Algorithmen geschehen, die im Folgenden erläutert werden.<sup>13</sup>

#### **Lineare Regression**

Die lineare Regression ist der einfachste Algorithmus. Er modelliert eine lineare Beziehung zwischen Eingabefeatures und Ausgabevariablen. Dieses Modell

<sup>12</sup> vgl. Patel 2020, 5f.

<sup>13</sup>

liefert nur ein aussagekräftiges Ergebnis, falls die verwendeten Daten eine wahre lineare Beziehung aufweisen, ansonsten liefert das Modell eine Unteranpassung der Daten. Durch die Einfachheit des Modells lässt sich das Ergebnis sehr leicht interpretieren, dies ist wichtig für die Betrachtung des angewandten maschinellen Lernens. Nicht nur technisches Fachpersonal, sondern auch nicht technisches Personal müssen mit der Lösung arbeiten und diese umsetzen können.<sup>14</sup>

### **Logistische Regression**

Ein anderer Klassifizierungsalgorithmus ist die logistische Regression. Die logistische Regression ist ebenfalls ein lineares Modell, jedoch trifft dieser Algorithmus seine Entscheidungen anhand der Klassenwahrscheinlichkeit. Die Klassenwahrscheinlichkeit ist die Wahrscheinlichkeit mit der der Datensatz in eine bestimmte Klasse passt. Die Stärken dieses Algorithmus sind dieselben wie des Linearen Regressionsalgorithmus. Er ist einfach zu interpretieren und ist bei vorwiegend nicht überlappenden Daten der passende Algorithmus.<sup>15</sup>

### **Softmax-Regression**

Die Softmax-Regression ist eine Bündelung von mehreren logistischen Regressionen. Die Verwendung zielt drauf aus, mehrere Klassen voneinander zu unterscheiden. Dabei werden mehrere logistische Regressionen durchgeführt, die die Wahrscheinlichkeiten der jeweiligen Klassen im Vergleich zu den anderen Klassen berechnet und auf eins nominiert. Die Gesamtwahrscheinlichkeit über alle Klassen ergibt somit eins. Das Softmax-Verfahren wird auch als Trennverfahren bezeichnet, da zur Unterscheidung der verschiedenen Klassen eine Gerade durch den Raum gezogen werden kann. Wie in Abbildung 2-3 dargestellt, zeigen die abgetrennten Bereiche an, welche Daten in welche Klasse eingeordnet werden.<sup>16</sup>

---

<sup>14</sup> vgl. Patel 2020, S. 12.

<sup>15</sup> vgl. Hirschle 2021, 54f.

<sup>16</sup> vgl. Hirschle 2021, 63f.

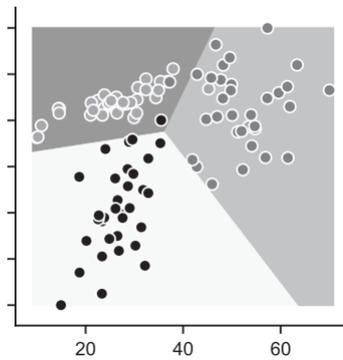


Abbildung 2-3 Darstellung lineare Trennung (Hirschle 2021, S. 63)

### 2.2.1.2. Klassifikation

Als Klassifikatoren werden Algorithmen bezeichnet, die Merkmalsvektoren auf kategoriale Zielvariablen abbilden. Im Folgenden werden drei Algorithmen dieser Art vorgestellt.<sup>17</sup>

#### Nearest Neighbour-Methoden

Nearest Neighbour-Methoden sind sehr einfache Algorithmen die zu den faule Lernende (Lazy Learners). Die faule Lernende Algorithmen werden auch instanzbasiertes Lernen oder nicht parametrische Methode genannt. Sie ordnen Daten zu Labels, indem sie die Daten anhand des Abstands zu bereits vorhandenen Punkten untersuchen und diese dem Label mit dem kürzesten Abstand zum nächsten Punkt einordnen. In Abbildung 2-4 ist ein Beispiel mit positiven und negativen Daten. Der schwarz hervorgehobene Punkt wird neu zugeordnet. Sein nächster Nachbar ist ein negativer Punkt, deshalb wird der neue Punkt negativ klassifiziert.

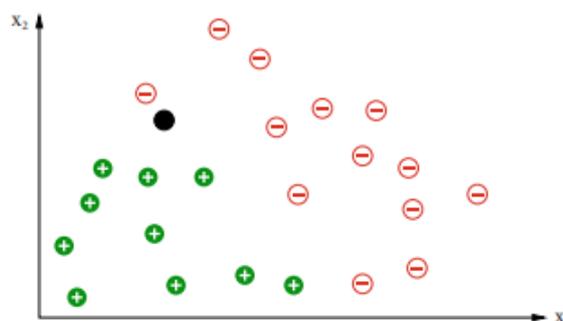


Abbildung 2-4 Beispiel der Nearest Neighbour-Methode (Ertel 2021, S. 218)

<sup>17</sup> vgl. Papp et al. 2022, S. 226.

Bei Betrachtung der Abbildung 2-5 können Schwachstellen der nächste Nachbar Methode erkannt werden. Wie in Abbildung 2-4 sind auch hier positive und negative Testdaten zu sehen. Der nächste Nachbar zum schwarz hervorgehoben Punkt ist in dieser Abbildung jedoch ein positiver Punkt. Demzufolge würde der schwarze Punkt nun als falsch positiv klassifiziert werden. Bei Betrachtung der gesamten Punktemenge fällt auf, dass es sich bei dem positiven nächsten Nachbar um einen Datenausreißer oder um einen falsch klassifizierten Punkt handeln könnte. Um diesen Fehler zu vermeiden kann die k-nächste Nachbar-Methode angewendet werden.<sup>18</sup>

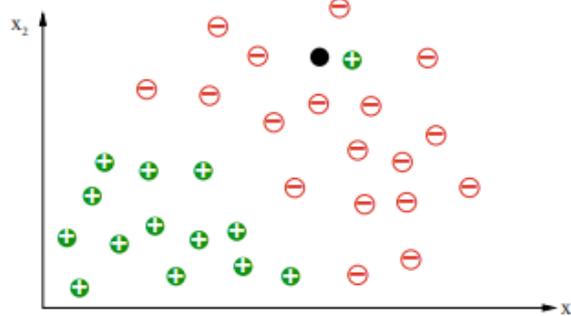


Abbildung 2-5 Beispiel k-Nearest Neighbour-Methode für k-nächste Nachbarn (Ertel 2021, S. 220)

Um neue Punkte einem Label zuzuordnen verwendet k-nächste Nachbarn (k-NN) eine Anzahl  $k$  ( $k$ = ein ganzzahliger Wert) der nächsten Punkte. Diese Punkte stimmen über die Benennung des neuen Punktes ab. Dadurch würde der schwarze Punkt in Abbildung 2.2-4 durch die mengenmäßig höher vorhanden negativen Punkte auch als negativ klassifiziert werden. Standardmäßig wird zur Messung des Abstands der euklidische Abstand verwendet. Die variable  $k$  kann frei gewählt werden. Jedoch kann bei einem sehr kleinen  $k$  der Algorithmus sehr flexibel werden, sodass er sich den Daten zu sehr anpasst. Bei einem zu großen  $k$  wird der Algorithmus zu unflexibel und zieht zu starren Grenzen.

Eine Stärke des k-NN Algorithmus ist es, dass er gegenüber von linearen Methoden sehr flexibel ist. Zudem ist er einfach und leicht interpretierbar. Die Schwäche des Algorithmus ist, dass bei hoher Anzahl an Daten der Algorithmus rechentechnisch ineffizient wird. Der k-NN Algorithmus wird bei Empfehlungssystemen angewendet. Die bekanntesten Beispiele sind Streaming Dienste für Filme und Musik. Diese schlagen dem Nutzer neue Inhalte vor, die ähnlich zu ihrer Auswahl sind und auch von anderen Nutzern mit gleichen

<sup>18</sup> vgl. Ertel 2021, S. 217–221.

Interessen genutzt werden. Dasselbe passiert bei Shopping Plattformen wie Amazon. Hier werden Nutzern ebenfalls Dinge vorgeschlagen die sie mögen (kollaboratives Filtern) und Dinge die sie in der Vergangenheit interessant fanden (inhaltsbasiertes Filtern).<sup>19</sup>

### **Baumbasierte Methoden**

Entscheidungsbäume basieren auf einer Reihe von Entscheidungen. Jede Entscheidung bildet ein Merkmal der Daten ab, wie die Verzweigungen eines Baumes. Am Ende des Entscheidungsbaum wird ein Blatt ohne Verzweigungen erreicht. Dieses Blatt gibt dem Beispiel ein Klassenlabel.

Merkmale können kategorische oder skalare Eigenschaften besitzen. Merkmale die in Kategorien unterteilt werden, sind beispielsweise die Farbe eines Autos (blau, grün, rot). Skalare Merkmale sind beispielsweise die Entfernung zum nächsten Bahnhof ( $<10$  km,  $\geq 10$  km). Bei skalaren Merkmalen wird ein Schwellenwert zur Einordnung des Beispiels verwendet. Je nachdem, ob das Beispiel größer und kleiner des Schwellenwerts ist, wird es in die jeweilige Richtung eingeteilt. Beim Anlernen des Entscheidungsbaum, werden jedes Mal neue Verzweigungen gebildet, solange bis in jedem Blatt nur noch Beispiele mit dem gleichen Label sind.<sup>20</sup>

Der größte Vorteil von Entscheidungsbäumen ist, dass durch ein einfaches, aber effizientes Verfahren, Wissen generiert wird. Durch die einfache Darstellung als Entscheidungsbaum, wie in Abbildung 2-6 dargestellt, können diese Informationen von Menschen leicht interpretiert und verstanden werden.<sup>21</sup>

---

<sup>19</sup> vgl. Patel 2020, 13f.

<sup>20</sup> vgl. Papp et al. 2022, 228f.

<sup>21</sup> vgl. Ertel 2021, 231f.

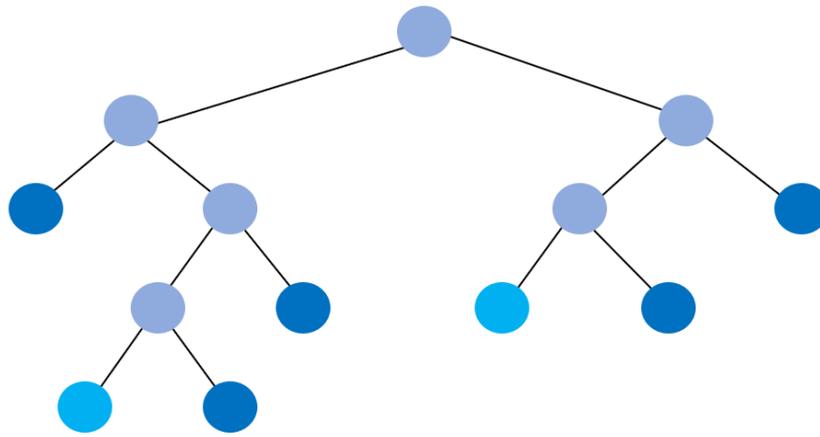


Abbildung 2-6 Entscheidungsbaum (In Anlehnung an Papp et al. 2022, 229)

Entscheidungsbäume sind allerdings anfällig für Überanpassungen. Durch die immer bessere Trennung der Daten, kann diese eventuell nicht auf neue Daten übertragen werden. Dies kann verhindert werden durch „Pruning“, einer Beschneidung des Baumes. Durch ein Kreuzvalidierungsverfahren kann der Baum nach Beschneidung auf seinen Erfolg getestet werden.<sup>22</sup>

### Support Vector Machine

Support Vector Machines (SVM) klassifizieren neue Punkte genau wie die k-NN-Methode nicht anhand aller bekannten Daten, sondern an sogenannten Support-Vektoren. Die Daten werden durch eine Klassengrenze getrennt. Support-Vektoren liegen nah an dieser Grenze und bilden den maximalen Abstand zur Trennfläche (siehe Abbildung 2-7).

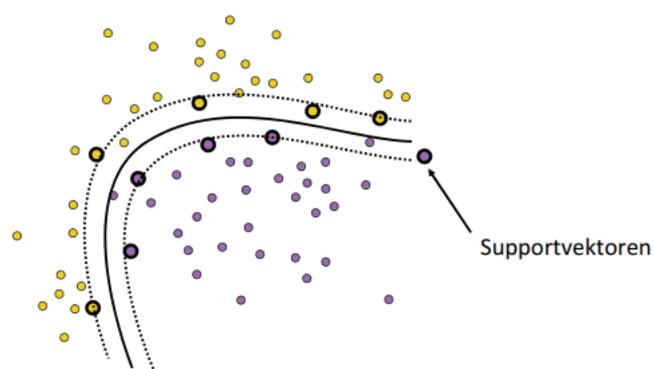


Abbildung 2-7 Beispiel für Support Vector Machine mit Klassengrenze (Papp et al. 2022, S. 227)

Sie definieren mit ihrem Einflussbereich die Klassengrenze. Neue Punkte können anhand der Koordinaten der Support-Vektoren bestimmt werden. Eine Schwäche

<sup>22</sup> vgl. Papp et al. 2022, 228f.

von SVM's ist das störende Merkmale die Methode beeinträchtigen, sodass die Auswahl der Merkmale besonders wichtig ist.<sup>23</sup>

## 2.2.2. Unüberwachtes Lernen

### 2.2.2.1. Clustering

Clustering-Verfahren sortieren und kategorisieren Daten in zusammenhängende Cluster (Gruppen). Die Einteilung in einem Cluster erfolgt für Daten mit ähnlichen Eigenschaften. Die Einteilung unterscheidet sich nach Auswahl der Algorithmen, die zum Clustern genutzt werden.<sup>24</sup>

#### **kMeans Algorithmus**

Bei einer bekannten Anzahl an Cluster wird der kMeans-Algorithmus angewendet. Dieser weist durch die Eingabe der Cluster in k, jeder Beobachtung genau einem zufälligen Cluster zu.<sup>25</sup> Diese Cluster werden durch ihre Mittelpunkte definiert. In nächsten Schritt werden die Abstände zwischen den Clustern und ihren zugeordneten Beobachtungen berechnet. Durch das Wiederholen der Klassifikation aller Daten zum nächsten Mittelpunkt und der Neuberechnung der Clustermittelpunkte werden die Gruppen optimiert.<sup>26</sup> Der Abstand der Beobachtungen zum Mittelpunkt wird mithilfe des quadratischen euklidischen Abstands berechnet, siehe folgende Formel:

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$$

Optimierungsprobleme lassen sich mithilfe des quadrierten euklidischen Abstandes einfacher berechnen, da es sich dabei um eine konvexe Funktion handelt. Beim kMeans-Algorithmus wird die Summe der quadrierten Abstände zwischen Mittelpunkten und zugeordneten Punkten mit folgender Formel berechnet:

$$d^2(p, q) = (q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2$$

---

<sup>23</sup> vgl. Papp et al. 2022, S. 227.

<sup>24</sup> vgl. Nguyen und Zeigermann 2021, S. 5.

<sup>25</sup> vgl. Ertel 2021, 262f.

<sup>26</sup> vgl. Ng und Soo 2018, S. 25–27.

Das Ziel des Algorithmus ist es diese Summe zu minimieren. Der Algorithmus optimiert so die Mittelpunkte mit den minimalsten Summen und identifiziert k-Gruppen mit entsprechenden Beobachtungen.<sup>27</sup>

## Hierarchisches Clustering

Im hierarchischen Clustering muss am Anfang keine Anzahl der Cluster festgelegt werden. Die Anzahl wird am Ende gewählt. Der Algorithmus erstellt aus den Beobachtungen ein Dendrogramm. Das Dendrogramm sieht aus wie ein umgedrehtes Baumdiagramm (siehe Abbildung 2-8).

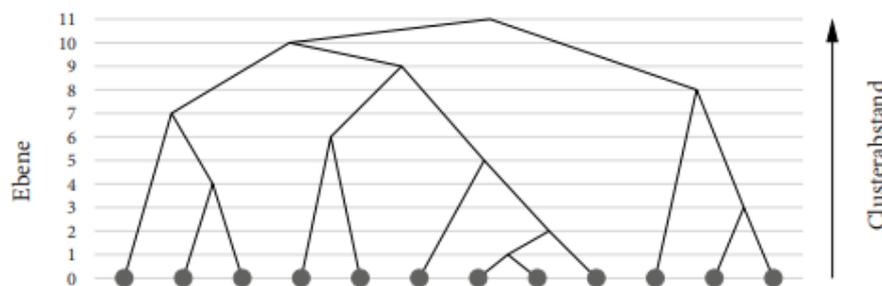


Abbildung 2-8 Darstellung eines Dendrogramms (Ertel 2021, S. 264)

Die Blätter des Dendrogramms sind einzelne Instanzen. Durch den Algorithmus werden ähnliche Daten zuerst miteinander verbunden bis alle Daten zu einem Stamm zusammengebunden sind. Durch das Beschneiden des Baumes kann bestimmt werden, wie viele Gruppen aus den Daten entstehen. Je höher beschnitten wird, desto weniger Gruppen werden generiert und umgedreht.<sup>28</sup>

### 2.2.2.2. Hauptkomponentenanalyse

Die Hauptkomponentenanalyse (PCA, principal component analysis) wird genutzt um Dimensionen (Hauptkomponenten) einer Datenmenge zu identifizieren und reduzieren.<sup>29</sup> Dies geschieht anhand eines mehrdimensionalen Raums. Dieser wird durch Achsen getrennt, wobei die Achsen nach Variablen bezeichnet werden. Die PCA rotiert das Koordinatensystem so lange bis ein neues Koordinatensystem mit neuen Achsen entsteht.

Die Grundlage dieser Umstrukturierung sind Kovarianzen zwischen den Daten. Die neuen Achsen enthalten nicht-redundante Informationsteile der Daten und

<sup>27</sup> vgl. Papp et al. 2022, 316f.

<sup>28</sup> vgl. Ertel 2021, S. 264.

<sup>29</sup> vgl. Ng und Soo 2018, S. 31–34.

sind so organisiert, dass die erste Dimension die höchsten Anteile an Varianz der Daten, die zweite Dimension den zweithöchsten Anteil an Varianz und so weiter enthalten. Bei einer Aufsummierung aller Daten, werden die Informationen wieder vollständig hergestellt. Durch die Dimensionsreduzierung werden nicht alle Daten verwendet, trotzdem bleiben mit der Hälfte der Variablen ca. 90% der Informationen aus den Daten erhalten.<sup>30</sup>

Ein Nachteil der Dimensionsreduktion mit der PCA-Methode ist die Annahme, dass die Dimension mit den größten Anteilen an Varianz auch die größte Aussagekraft besitzt. Das ist nicht immer der Fall. Ein weiterer Nachteil ist die Erzeugung einer orthogonalen Dimension. Das bedeutet das Dimensionen senkrecht aufeinander stehen. Jedoch müssen diese nicht immer senkrecht zueinanderstehen. Um das zu umgehen, kann eine Unabhängigkeitsanalyse (ICA, independent component analysis) durchgeführt werden. Die ICA lässt nicht-orthogonale Dimensionen zu. Diese dürfen sich für die Analyse allerdings nicht überlappen. Um Ergebnisse der PCA zu verifizieren kann im Nachhinein immer eine ICA gemacht werden.<sup>31</sup>

### 3. Datenerhebung

Um herauszufinden, welche Einflussfaktoren das Täuschungsverhalten von Studierende beeinflussen, wurde eine quantitative und qualitative Studie durchgeführt.

Für diese Arbeit wurde eine Umfrage erstellt, die Informationen über die privaten und schulischen Tätigkeiten der Studierenden erfasst. Zudem wurde erfasst, welche verschiedenen Schummel Methoden die Studierenden kennen und bei welcher Fragestellung und Prüfungsform sie diese anwenden würden.

Die Umfrage wurde online über einen E-Mail-Verteiler an sämtliche Studierende der Hochschule Neu-Ulm verteilt. Zudem wurde in vier Vorlesungen von Professorin Dr. Sonja Köppl die Umfrage vorgestellt und verteilt, indem die Studierenden extra Zeit bekommen haben, um an der Umfrage teilnehmen zu

---

<sup>30</sup> vgl. Hirschle 2021, S. 249.

<sup>31</sup> vgl. Ng und Soo 2018, S. 40–43.

können. Am 06.12.2022 wurde die Umfrage veröffentlicht und war bis zum 02.01.2023 online verfügbar.

Die Umfrage umfasst 42 Fragen aufgeteilt in 5 Gruppen und die Bearbeitung dauerte ca. 12 Minuten. Die Gruppen wurden wie folgt aufgeteilt:

- Teil A: Allgemeine Fragen zum Studium
- Teil B: Allgemeine Fragen zum Privatleben
- Teil C: Fragen zu Prüfungen
- Teil D: Fragen zum Schummeln
- Teil E: Demographische Fragen

Im Fragenteil A wurden Fragen zum Studiengang, Semester und Notendurchschnitt gestellt. Im nächsten Fragenteil wurde die Zufriedenheit mit dem Studium und der Hochschule, die persönliche Motivation und der persönliche Leistungsdruck abgefragt. Zudem gab es Fragen zur Vorbereitung der Vorlesung. Desweiteren wurden Freizeitaktivitäten, Interessen, Nebenjob, ehrenamtliche Tätigkeiten, Social-Media-Verhalten, familiäre Verpflichtungen und die Religiosität abgefragt.

Im dritten Teil des Fragebogens wurden Fragen zur Teilnahme an online Prüfungen gestellt. Zudem wurden die benötigte Ausstattung und der Vergleich zwischen Präsenz und Online Prüfungen in Bezug auf Wohlfühlen, Fairness und Besseres abschneiden abgefragt.

Teil D des Fragebogens beschäftigte sich mit Fragen zur Einstellung zum Schummeln, den Konsequenzen beim Schummeln, bekannten Schummel Methoden, der Einfluss des Dozenten auf das Schummel Verhalten und die Anwendung von Schummel-Methoden in Prüfungen und Aufgabentypen. Im letzten Teil des Fragebogens wurden die demographischen Daten wie Alter, Geschlecht und die Wohnsituation abgefragt.

Der vollständige Fragebogen befindet sich im Anhang unter 1. Umfrage. Insgesamt haben 460 Studierende die Umfrage bearbeitet, wovon 263 die Umfrage vollständig abgeschlossen haben.

## 4. Auswertung mit Python

Für die Auswertung wurden nur die vollständig ausgefüllten Fragebögen berücksichtigt. Zudem wurden die Ausreißer, Raser, Straightling und unsinniges Feedback in offenen Fragen aus den Fragebögen aussortiert. Zur Analyse wurden 257 vollständige Fragebögen herangezogen.

### 4.1. Aufbereitung der Daten

Zur besseren Auswertung wurden die Ergebnisse in Excel auf unterschiedliche Blätter aufgeteilt. Die Aufteilung erfolgte nach:

- Demographische Daten
- Zufriedenheit
- Private Aktivitäten
- Hobbies
- Interessengebiete
- Familiäre Verpflichtungen
- Sozial Media
- Schulische Aktivitäten
- Prüfungsverhalten
- Digitale Prüfungen Teilnahme
- Digitale Prüfungen Schummeln
- Schummel-Verhalten
- Versuch
- Einfluss Dozent
- Bekannte Schummel-Methoden
- Matrix (Schummel-Methode Anwendung, Schummel-Methode Aufgabentyp)

Bei der Aufteilung wurden bei offenen Antworten Möglichkeiten, falls es Sinn ergeben hat, Antworten in Kategorien eingeteilt oder eine neue Antwortmöglichkeit anstatt „Sonstiges“ geschaffen. Bei der Frage nach dem Studiengang wurde der Studiengang Informationsmanagement und Unternehmenskommunikation als Antwortmöglichkeit hinzugefügt, da 29 Teilnehmer diesen Studiengang angegeben hatten.

Die Frage nach dem Nebenjob ist eine offene Frage. Hier wurden nach den Antworten der Teilnehmer Kategorien erstellt und alle ähnlichen Antworten einer dieser Kategorien zugeordnet. Beispielsweise wurden alle Werkstudenten Nebenjobs in die Kategorie „Werkstudent“ eingeteilt, egal in welcher Branche sie tätig sind. Ebenfalls wurden alle Nebenjobs mit den Angaben Kellner/-in, Servicekraft und Küchenhilfe in der Kategorie „Gastronomie“ zusammengefasst. Endstanden sind folgende Kategorien:

- Handwerk
- Verkauf

- Verwaltung
- studentische Hilfskraft
- Beratung
- medizinische Aushilfe
- Vollzeit
- bezahlte Vereinstätigkeit
- Nebenjob
- Nachhilfe
- Produktion
- Lager
- Vertrieb
- Werkstudent
- Gastronomie

Als weiterer Schritt der Datenaufbereitung wurden alle Zeitangaben, welche in Intervallen abgefragt wurden, als Mittelwert des jeweiligen Intervalls ausgegeben. Dies erfolgte für eine bessere Berechnung und Auswertung. Ein Beispiel hierfür ist die Angabe 0-5 Stunden. Diese wurde in 2,5 Stunden geändert. Diese Aufbereitung ist bei den Arbeitsstunden, den Stunden fürs Ehrenamt, die Zeit für Hobbies/Sport, die Zeit für Social Media und die Vor- und Nachbearbeitungszeit für die Vorlesung. In der nachfolgenden Tabelle 4-1 befindet sich die gesamte Aufschlüsselung.

Angabe in Umfrage	Ersatz	Angabe in Umfrage	Ersatz
0-5	2,5	0-3	1,5
6-10	7,5	4-7	5,5
11-15	12,5	8-11	9,5
16-20	17,5	12-15	13,5
21-25	22,5	16-19	17,5
26-30	27,5	20-23	21,5
31-35	32,5	24-27	25,5
36-40	37,5	28-31	29,5
		32-35	33,5
		36-39	37,5
		40 und mehr	41,5

*Tabelle 4-1 Aufschlüsselung der Ersetzungen der Zeitangaben*

Die Antworten der Frage nach dem Notendurchschnitt wurden, wie die Zeitangaben, zur einfacheren Berechnung und Auswertung in Mittelwerte geändert. In Tabelle 4-2 befindet sich die vollständige Aufschlüsselung.

Noten	Ersatz
1,0-1,5	1,25
1,6-2,0	1,75
2,1-2,5	2,25
2,6-3,0	2,75
3,1-3,5	3,25
3,6-4,0	3,75

Tabelle 4-2 Aufschlüsselung der Notenersetzung

Da bei der Frage nach dem Notendurchschnitt eine hohe Anzahl der Teilnehmer diesen nicht angegeben haben, wurde der Mittelwert von 2,141753 für sie eingetragen. Die Berechnung des Mittelwerts wurde in Python mithilfe der „describe()“-Funktion aus der matplotlib Bibliothek (siehe Abbildung 4-1). Die Funktion „describe()“ berechnet für alle numerischen Daten in der eingelesenen Tabelle die Anzahl der Antworten, den Mittelwert (M), die Standardabweichung (SD), das Minimum (min), das Maximum (max) das 25%, 50% und 75% Quartil. Die Funktion wird im weiteren Verlauf häufiger verwendet.

	count	mean	std	min	25%	50%	75%	max
<b>Fachsemester</b>	255.0	4.015686	2.517605	1.00	2.00	4.000000	6.00	11.00
<b>Notendurchschnitt</b>	257.0	2.141753	0.485037	1.25	1.75	2.141753	2.25	3.75
<b>Alter</b>	238.0	23.323529	4.822630	18.00	21.00	22.000000	24.00	56.00

Abbildung 4-1 Ausgabe der "describe()"-Funktion aus Python (Eigene Berechnung)

Im nachfolgenden Unterkapitel werden verschiedene komplexere Python Codes zum Verständnis der Auswertung aufgezeigt und erläutert.

## 4.2. Algorithmen und Methoden in Python zur Auswertung

Um ein Clustering Verfahren in Python durchführen zu können, müssen nicht numerische Daten aufbereitet werden. Algorithmen des maschinellen Lernens können nur mit numerischen Daten arbeiten. Um keinen mathematischen Zusammenhang zwischen den Daten herzustellen, sollten diese in „0“ und „1“ geändert werden. Zudem dürfen keine leeren Zellen im DataFrame existieren.

Um diesen Zustand der Daten herzustellen, wurden die leeren Zellen mit dem Wert „na“ befüllt. Zudem wurden alle Ja's in den Wert „1“ und alle Neins in den

Wert „0“ umgewandelt. In der Spalte Alter und Fachsemester wurden die leeren, durch „na“ ersetzen Zellen durch den Wert „0“ ausgetauscht.

#### 4.2.1. LabelEncoder und OneHotEncoder

Nach dieser Vorbereitung wurden alle übrigen Spalten, die sting (wörtliche) Werte enthalten mithilfe des LabelEncoder und des OneHotEncoder in numerische Werte geändert. In Abbildung 4-2 ist der benutzte Code dargestellt. Im Code wurde eine Schleife aufgebaut, die die Werte in den Spalten durch das LabelEncoder in Zahlenwerte umwandelt. Danach werden für jeden der Werte im LabelEncoder eine neue Spalte angelegt mit den Werten „1“ dafür, dass der Wert vorhanden ist und „0“ für den Fall, dass der Wert nicht im DataFrame enthalten ist. Hierbei wird durch eine weite Schleife überprüft, ob bestimmte Spalten enthalten sind, falls diese enthalten sind, werden sie zusätzlich mit dem ursprünglichen Spalten-namen versehen. Die neuen Spalten werden an den DataFrame angehängt und die umgecodete Spalte wird entfernt. Des Weiteren werden die Spalten „Sonstiges“ und „na“ durch eine weitere Schleife entfernt, falls sie vorhanden sind. Dies geschieht, da leere Werte und „Sonstige“ Werte wenig aussagekräftig sind und die Auswertung verfälschen könnte.

```
In [4]: one_hot_encoding_columns = ['Schummelhaltung', 'Studiengang', 'Geschlecht', 'Wohnort', 'Zufriedenheit_Studium',
                                   'Zufriedenheit_Hochschule', 'Motivation', 'Leistungsdruck']

for x in one_hot_encoding_columns:
    cluster[x] = labelencoder.fit_transform(cluster[x])
    encoder_cluster = pd.DataFrame(encoder.fit_transform(cluster[[x]]).toarray())

    classes = labelencoder.classes_
    desired_classes = ["Stimme voll und ganz zu", "Stimme zu", "Weder noch", "Stimme nicht zu",
                      "Stimme überhaupt nicht zu"]
    class_names = []
    for class_ in classes:
        if class_ in desired_classes:
            class_names.append(x + "_" + class_.replace(" ", "_"))
        else:
            class_names.append(class_)

    encoder_cluster.columns = class_names
    cluster = cluster.join(encoder_cluster)

columns_to_drop = ["Sonstiges", "na"]
for column in columns_to_drop:
    if column in cluster.columns:
        cluster.drop(column, axis=1, inplace=True)

cluster.drop([x], axis=1, inplace = True)

cluster.head()
```

Out[4]:

	Schummeln_TH_Abgabe	Schummeln_TH_Test	Schummeln_Pool_Prüfung	Schummeln_Schriftliche_Zoom	Schummeln_Mündliche_Zoom	Fachsemester	Not
0	1	0	0	0	0	3.0	
1	0	0	0	0	0	1.0	
2	0	0	0	0	0	1.0	
3	0	0	0	0	0	1.0	
4	0	0	0	0	0	1.0	

5 rows x 72 columns

Abbildung 4-2 OneHotEncoding der Spalten mit string Werten (Eigene Darstellung)

## 4.2.2. StandardScaler

Um eine Vergleichbarkeit der Werte sicherzustellen, werden diese normalisiert. Das bedeutet die Daten werden auf eine ähnliche Skala gebracht um mögliche Ausreißer oder unterschiedliche Skalierungen von Spalten zu minimieren. Dies ist in Abbildung 4-3 als Code dargestellt. Mithilfe der „StandardScaler()“-Funktion und der „fit\_transform()“-Methode werden Mittelwerte und Standardabweichungen für jede Spalte berechnet und dazu verwendet, die Daten zu normalisieren.

```
In [10]: x1 = cluster.loc[:, :].values
x_stand1 = StandardScaler().fit_transform(x1)
X_stand1 = pd.DataFrame(data = x_stand1, columns = cluster.columns)
X_stand1.head()
```

```
Out[10]:
```

	Fachsemester	Notendurchschnitt	Alter	Fußball	Basketball	Tennis	Tanzen	Yoga	Schwimmen	Fitnessstudio	...	Motiviert	Sehr motiviert
0	-0.389471	2.289330e+00	0.182844	-0.377124	-0.201211	-0.240027	-0.349672	-0.274434	-0.305969	1.203170	...	-0.943242	-0.230822
1	-1.180730	6.591016e-07	-0.339278	-0.377124	-0.201211	-0.240027	-0.349672	-0.274434	-0.305969	-0.831137	...	1.060173	-0.230822
2	-1.180730	6.591016e-07	-0.469809	2.651650	4.969909	-0.240027	-0.349672	-0.274434	-0.305969	1.203170	...	1.060173	-0.230822
3	-1.180730	6.591016e-07	-0.339278	-0.377124	-0.201211	4.166190	-0.349672	-0.274434	-0.305969	-0.831137	...	-0.943242	4.332347
4	-1.180730	2.289330e+00	0.182844	-0.377124	-0.201211	-0.240027	-0.349672	-0.274434	-0.305969	1.203170	...	1.060173	-0.230822

5 rows x 63 columns

Abbildung 4-3 Normalisierung des DataFrames (Eigene Darstellung)

## 4.2.3. PCA

Um diese große Datenmenge wieder zu komprimieren wird eine Dimensionsreduktion durchgeführt. Die PCA (siehe Abbildung 4-4) reduziert hochdimensionale Daten in niedrigere Dimensionen, ohne zu viele Informationen zu verlieren. Durch die Angabe der in „PCA(n\_components=2)“ werden, in diesem Fall, zwei Komponenten erstellt. Hier kann eine beliebige Zahl eingetragen werden.

```
In [12]: from sklearn.decomposition import PCA
pca_1 = PCA(n_components=2)
principalComponents = pca_1.fit_transform(X_stand1)

principalDf1 = pd.DataFrame(data = principalComponents, columns = ["principal component 1", "principal component 2"])
principalDf1.head()
```

```
Out[12]:
```

	principal component 1	principal component 2
0	-0.140814	-0.479671
1	-2.364996	-0.427549
2	-4.281976	-1.440439
3	-4.153961	1.197279
4	-2.256153	-0.841405

Abbildung 4-4 Code der PCA (Eigene Darstellung)

#### 4.2.4. KMeans

Der kMeans Algorithmus wird aus der sklearn Bibliothek geladen. Beim Einstellen des Algorithmus muss angegeben werden, wie viele Cluster der Algorithmus finden soll. In der Abbildung 4-5 werden 3 Cluster gebildet. Als nächster Schritt werden die Labels ausgegeben und in einem weiteren Schritt die Zentren der Cluster. Als zusätzlichen Schritt gibt Python die visuelle Darstellung der Einteilung der Cluster wieder. Die Cluster werden zur Auswertung der Daten an die ursprünglichen unveränderten Daten angehängt.

```
In [ ]: from sklearn.cluster import KMeans
# KMeans-Modell erstellen
km_model = KMeans(n_clusters=3, random_state=0).fit(principalDf1)

In [ ]: km_model.labels_

In [ ]: km_model.cluster_centers_

In [ ]: cluster_y = principalDf1.copy()
cluster_y["cluster"] = km_model.labels_
sns.relplot(x = "principal component 1", y = "principal component 2", hue = "cluster", data = cluster_y)
plt.show()

In [ ]: cluster_y = cluster.copy()
cluster_y["cluster"] = km_model.labels_
cluster_y.head()
```

Abbildung 4-5 Code des kMeans Algorithmus (Eigene Darstellung)

## 5. Interpretation der Ergebnisse

Bei der Analyse der Daten wurden sowohl eine deskriptive Auswertung als auch eine Clusteranalyse durchgeführt. In der deskriptiven Auswertung wurden Mittelwerte, Standardabweichungen und Häufigkeiten berechnet, um eine Übersicht über die Verteilung der Antworten zu erhalten. Die Clusteranalyse wurde genutzt, um Gruppen von Teilnehmern zu identifizieren, die ähnliche Antwortmuster aufwiesen, und diese Gruppen zu beschreiben.

### 5.1. Deskriptive Auswertung

Im ersten Teil der Auswertung wurden die demographischen Daten untersucht. Hierbei handelt es sich um Studiengang, Fachsemester, Notendurchschnitt, Alter, Geschlecht und Wohnort der Befragten.

In Hinblick auf die Verteilung der Studiengänge ist festzustellen, dass die meisten Teilnehmer (19,21%) vom Studiengang Betriebswirtschaft (BWL) sind.

Anschließend kommen die Studiengänge Wirtschaftsingenieurwesen (WIN) mit 15,68%, Wirtschaftspsychologie (WSP) mit 13,72%, Betriebswirtschaft im Gesundheitswesen (BGW) mit 13,33% und Informationsmanagement und Unternehmenskommunikation (IMUK) mit 11,37%. Die Vollständige Auflistung aller Antworten zum Studiengang befindet sich in Abbildung 5-1. In Tabelle 5-1 befindet sich die Auflistung aller Studiengänge mit der absoluten und prozentuellen Anzahl inklusive der Einteilung der Studiengänge in die Fakultäten der Hochschule Neu-Ulm. Auffällig ist das die meisten Studiengänge aus der Fakultät Wirtschaftswissenschaften (WW) kommen.

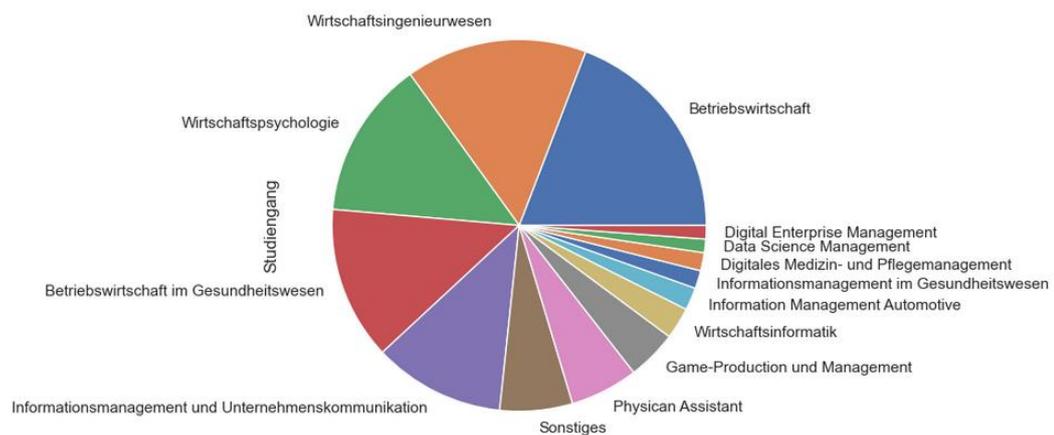


Abbildung 5-1 Verteilung der Studiengänge (Eigene Darstellung)

Studiengang	Anzahl der Befragten	Anzahl der Befragten in %	Fakultät
Betriebswirtschaft	49	19,21%	WW
Wirtschaftsingenieurwesen	40	15,68%	WW
Wirtschaftspsychologie	35	13,72%	WW
Betriebswirtschaft im Gesundheitswesen	34	13,33%	GM
Informationsmanagement und Unternehmenskommunikation	29	11,37%	IM
Sonstiges	16	6,27%	
Physican Assistant	15	5,88%	GM
Game-Production und Management	11	4,31%	IM
Wirtschaftsinformatik	7	2,74%	IM
Information Management Automotive	5	1,96%	IM
Informationsmanagement im Gesundheitswesen	4	1,56%	GM
Digitales Medizin- und Pflegemanagement	4	1,56%	GM
Data Science Management	3	1,17%	IM
Digital Enterprise Management	3	1,17%	WW

Tabelle 5-1 Tabelle der Studiengänge mit Fakultäten (Eigene Darstellung)

Das Alter variierte zwischen 18 und 56 Jahren, wobei ca. 30 % der Teilnehmer zum Zeitpunkt der Umfrage 22 und 23 Jahre als waren ( $M=23,32$ ;  $SD=4,82$ ). In Bezug auf das Fachsemester ist festzustellen, dass alle Semester vertreten waren. Der durchschnittliche Umfrageteilnehmer befand sich im 4. Fachsemester ( $SD=2,51$ ), die meisten Teilnehmer (23,52%) befanden sich im ersten Semester. Der Notendurchschnitt der Befragten lag durchschnittlich 2,14 mit einer Standardabweichung von 0,48. An der Umfrage nahmen 57,58% weibliche und 33,46 männliche Studierende teil (siehe Abbildung 5-2).

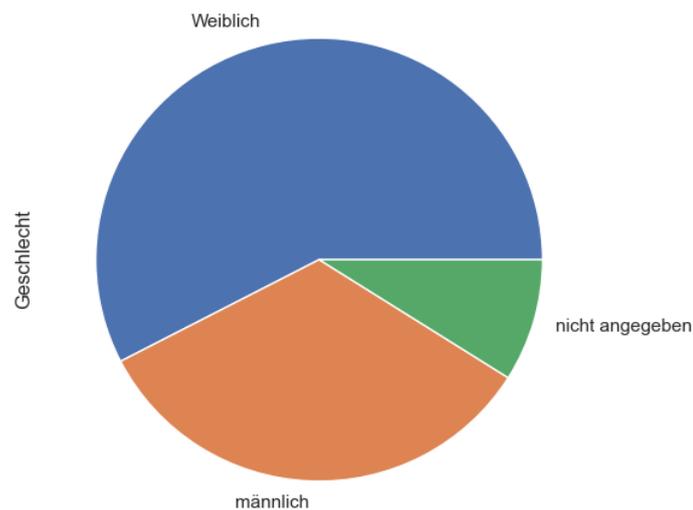


Abbildung 5-2 Geschlechtsverteilung der Umfrageteilnehmer (Eigene Darstellung)

Zu der Frage der Wohnsituation gaben 43,2% der Studenten an bei ihren Eltern zu wohnen, dargestellt in Abbildung 5-3). In einer eigenen Wohnung wohnten nach Angabe der Studierenden 23,2% und in einem WG-Zimmer 18,4%. Das Studentenwohnheim gaben 12,8% der Teilnehmer als Wohnort an.

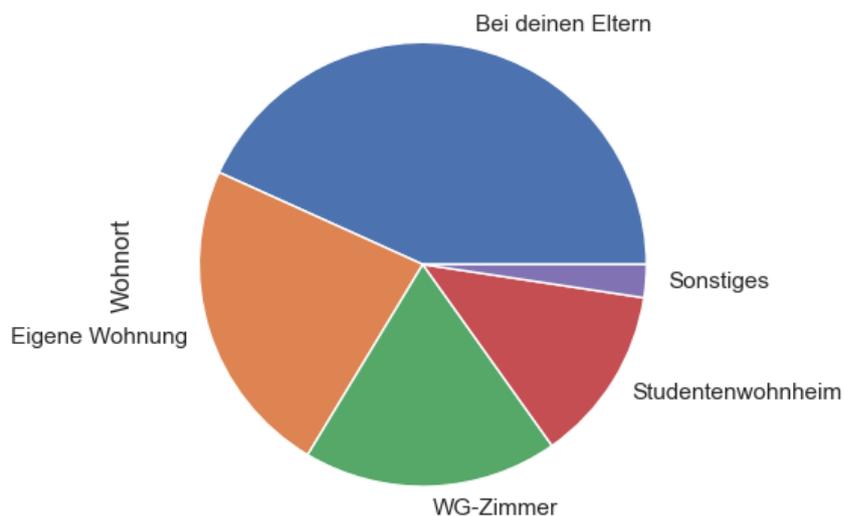


Abbildung 5-3 Wohnsituation der Umfrageteilnehmer (Eigene Darstellung)

Im nächsten Teil der Auswertung wurde die Zufriedenheit, der Leistungsdruck und die Motivation betrachtet. Dies wurde mithilfe einer Likert-Skala abgefragt. Eine Likert-Skala ist eine psychometrische Skala. Diese Skala repräsentiert die Zustimmung eines Merkmals anhand einer Ratingskala.<sup>32</sup> In der folgenden Abbildung 5-4, ist die Auswertung der Zufriedenheit mit dem Studium, der Hochschule abgebildet.

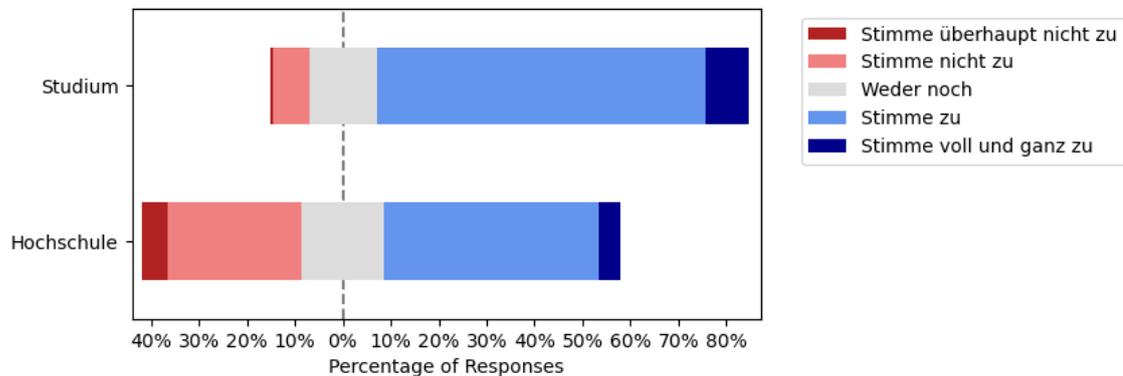


Abbildung 5-4 Auswertung der Zufriedenheit mit dem Studium und der Hochschule mittels Likert-Skala (Eigene Darstellung)

Bei der Auswertung ist zu erkennen, dass 77,73% der Teilnehmer an der Umfrage angeben haben, zufrieden mit ihrem Studium zu sein. In der Umfrage wählten 68,75% der Teilnehmer „Stimme zu“ und 8,98% der Teilnehmer „Stimme voll und ganz zu“. Nur insgesamt 8,2% der Teilnehmer sind nicht oder gar nicht zufrieden mit ihrem Studium. In Bezug auf die Zufriedenheit mit der Hochschule gaben ca. die Hälfte (49,4%) das Befragten an, das sie zufrieden sind. Es gaben 44,7% „Stimme zu“ und 4,7% „Stimme voll und ganz zu“ in der Umfrage an. Ein Drittel (33,33%) das Befragten gaben an, mit der Hochschule nicht und gar nicht zufrieden zu sein.

Zur Frage, ob die Teilnehmer der Umfrage einen Leistungsdruck verspüren, stimmen 22,83% das Befragten voll und ganz zu und 37,4% stimmten zu. Ein Fünftel (20,07%) der Befragten gab an keinen Leistungsdruck zu verspüren. Hinsichtlich der Motivation gaben 47,26% der Umfrage Teilnehmer an, motiviert im Studium zu sein. Die Ergebnisse sind in Abbildung 5-5 visuell dargestellt.

<sup>32</sup> vgl. o.V. o.J.

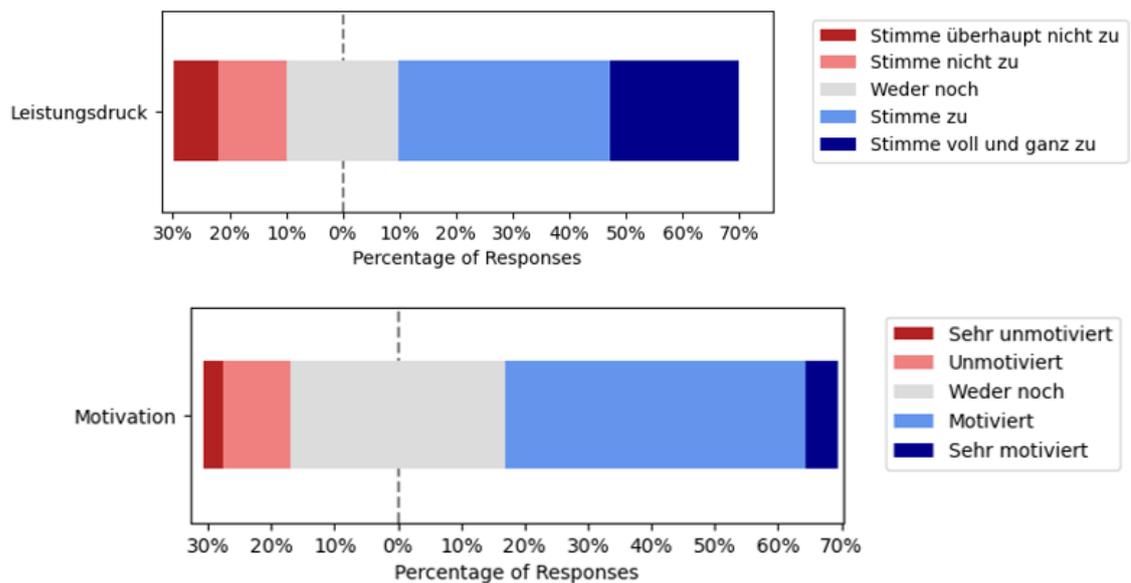


Abbildung 5-5 Auswertung des Leistungsdrucks und der Motivation der Befragten mittels Likert-Skala (Eigene Darstellung)

Im weiteren Teil der Auswertung werden die Hobbies betrachtet. Im Durchschnitt verbrachten die Befragten 10,26 h mit der Ausübung ihrer Hobbies und Sportarten, die Standardabweichung beträgt 8,40h. Die meisten Befragten gaben an das die Freunde Treffen (172 Befragte) in ihrer Freizeit. Zudem war das Fitnessstudio (105), das Lesen (90) und der Besuch einer Bar oder eines Clubs (89) sehr beliebte Freizeitbeschäftigungen. Die wenigsten Befragten spielten in ihrer Freizeit Poker (5), Handball (4) und Kampfsport (4). Die vollständige Auswertung der gesamten Freizeitaktivitäten befindet sich in Abbildung 5-6.

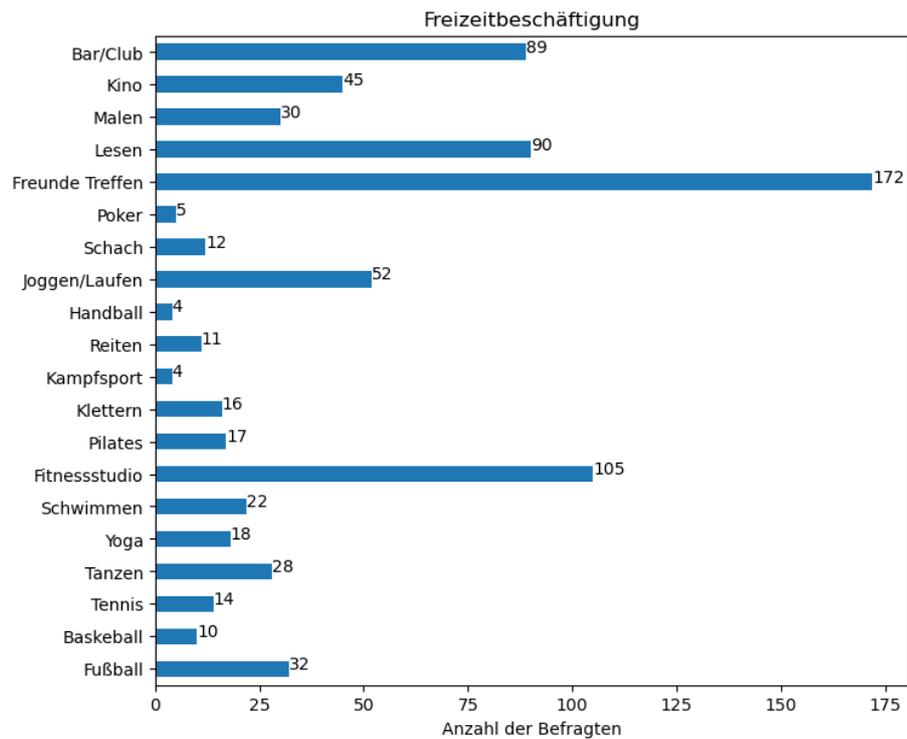


Abbildung 5-6 Freizeitbeschäftigung der Befragten (Eigene Darstellung)

Im folgenden Teil der Auswertung werden die schulischen Aktivitäten betrachtet. Die Befragten gaben an im Durchschnitt 12,31 Stunden in der Woche für die Vor- und Nachbearbeitung der Vorlesung zu brauchen. Die Standardabweichung hierfür beträgt 8,9 Stunden in der Woche. Die Frage nach dem regelmäßigen Besuch der Vorlesung wurde mithilfe einer vier-teiligen Likert-Skala abgefragt, dargestellt in Abbildung 5-7.

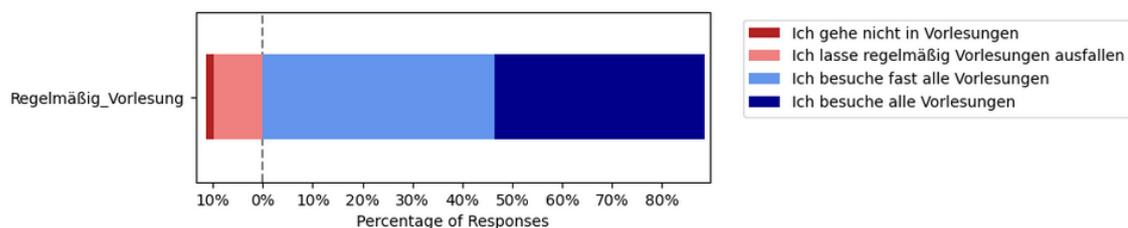


Abbildung 5-7 Regelmäßiger Besuch der Vorlesung (Eigene Darstellung)

Dabei gaben 46,45% der Befragten an, dass sie fast alle Vorlesungen besuchten und 42,12% der Befragten gaben an, dass sie alle Vorlesungen besuchten. Nur 1,57% der Befragten gaben an, keine Vorlesung zu besuchen.

Im letzten Teil der Auswertung werden die Verhaltensweisen in Prüfungen betrachtet. In der Umfrage gaben 78,80% der Teilnehmer an, dass die an einer Online Prüfung teilgenommen haben. Insgesamt 36,94% der Befragten gaben

an, das sie sich bei Präsenz-Prüfungen wohler fühlte (siehe Abbildung 5-8). Bei Online Prüfungen gaben nur 22,89% der Befragten an sich wohler zu fühlen.

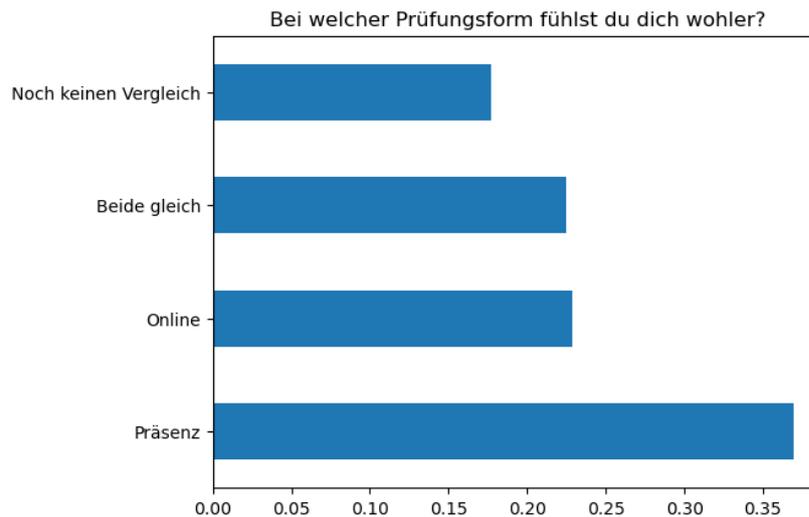


Abbildung 5-8 Angabe, bei welcher Prüfung sich die Befragten wohler fühlen (Eigene Darstellung)

Des Weiteren gaben 48,17% der Umfrage Teilnehmer an, die Präsenz Prüfung fairer zu finden (siehe Abbildung 5-9). Im Gegensatz dazu fanden 6,88% der Befragten eine Online Prüfung fairer.

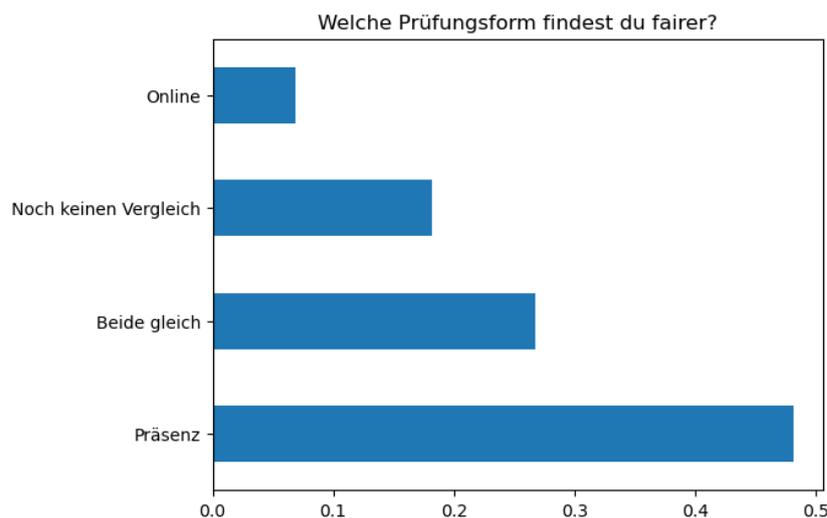


Abbildung 5-9 Angabe, welche Prüfung Befragte als fairer empfinden (Eigene Darstellung)

In der Umfrage gaben 141 Teilnehmer an bei einem Take Home Examen mit Moodle<sup>33</sup> Abgabe teilgenommen zu haben. 133 Teilnehmer gaben an, an einem Take Home Examen als Moodle Test teilgenommen zu haben. Die mündliche

<sup>33</sup> Das Kursmanagementsystem Moodle dient an Hochschulen dazu Vorlesungsmaterialien von Professoren mit Studierenden zu teilen und kann zusätzlich als Prüfungsplattform genutzt werden.

Prüfung über Zoom<sup>34</sup> wurde von 106 Befragten durchgeführt. Eine Prüfung im Pool EDV Raum wurde von 61 der Befragten durchgeführt. An einer Schriftlichen Prüfung über Zoom mit Kamera haben 52 Befragte teilgenommen, dargestellt in Abbildung 5-10.

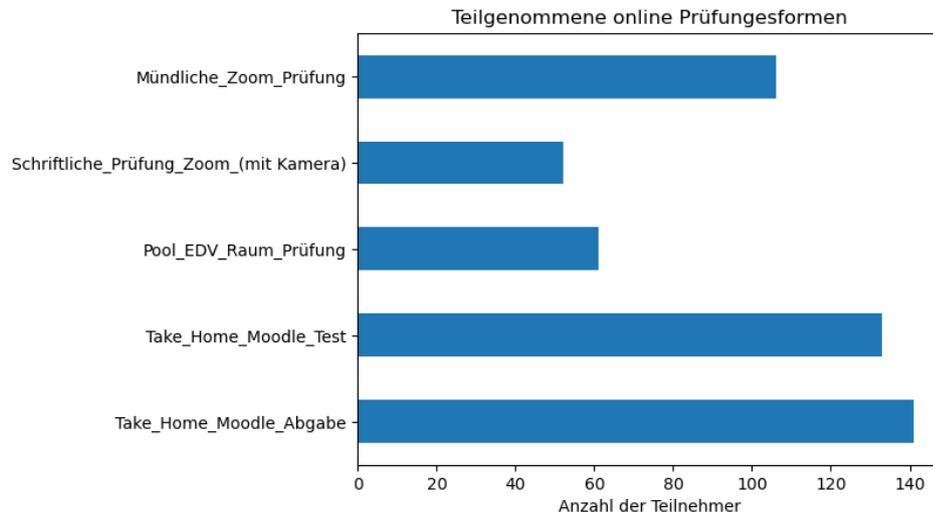


Abbildung 5-10 Angaben zur Teilnahme an online Prüfungsformen

Die Frage nach dem Schummeln in online Prüfungsformen beantworteten 60 Teilnehmer mit Ja für Take Home Examen als Moodle Abgabe und 62 bei Take Homes als Moodle Test. 18 Teilnehmer gaben an in Schriftlichen Prüfungen über Zoom mit Kamera geschummelt zu haben und 22 in Mündlichen Prüfungen über Zoom, dargestellt in Abbildung 5-11.

---

<sup>34</sup> Die Kommunikation Software Zoom dient an Hochschulen als online Tool um Vorlesungen und Prüfungen abzuhalten.

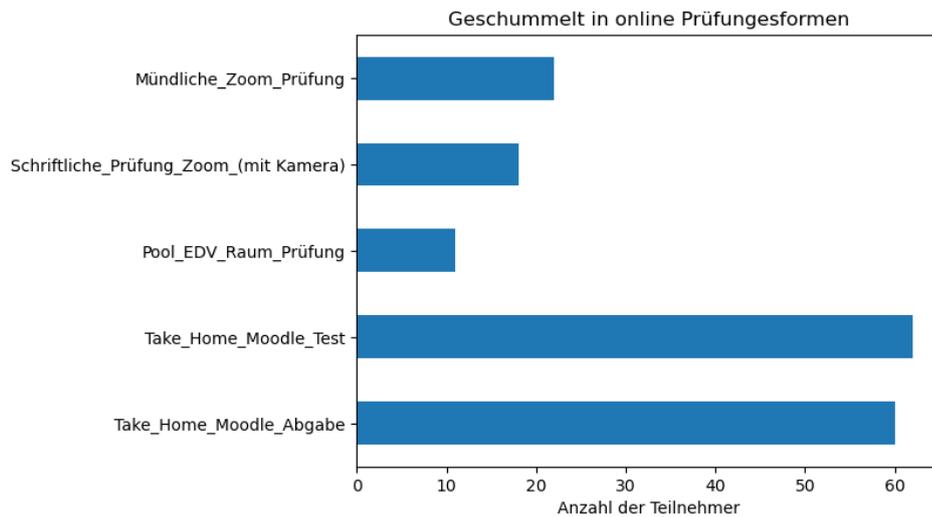


Abbildung 5-11 Angabe bei welchen Online Prüfungen geschummelt wurde (Eigene Darstellung)

Des Weiteren gaben 89,62% der Teilnehmer der Umfrage an, dass ihnen die Konsequenzen bekannt sind, falls sie beim Schummeln erwischt werden. Nur ca. einem Fünftel der Befragten sind diese Konsequenzen nicht bekannt.

In der Umfrage wurde die Haltung zum Schummeln in Online Prüfungen abgefragt (siehe Abbildung 5-12). Dabei gaben 59,10% der Befragten an, dass sie eine neutrale Haltung zum Schummeln haben. Insgesamt 13,35% finden Schummeln ok oder finden Schummeln ok, wenn Freunde schummeln. Verwerflich finden 21,05% der Befragten das Schummeln in Online Prüfungen.

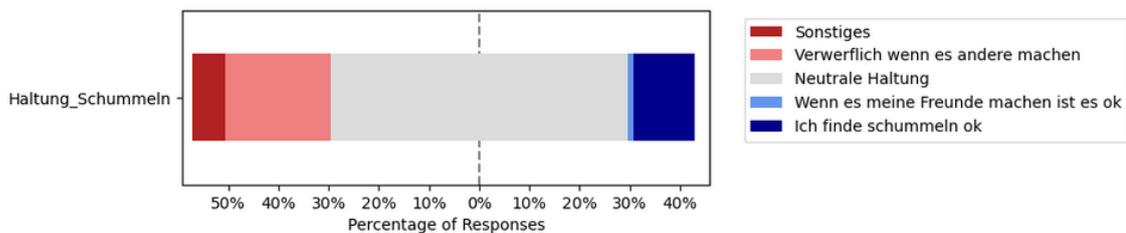


Abbildung 5-12 Haltung zum Schummeln bei Online Prüfungen (Eigene Darstellung)

Die Aufklärung der Konsequenzen vor den Prüfungen halten nach Angaben der Befragten nur 8,40% der Teilnehmer vom Schummeln ab. Die Exmatrikulation würde 32,77% der Befragten vom Schummeln abhalten. Das nicht bestehen der Prüfung (28,57%) und der Ausschluss von der Prüfung (15,54%) würde hinter der Exmatrikulation die meisten Befragten vom Schummeln abhalten. Eine vollständige Übersicht befindet sich in Abbildung 5-13.

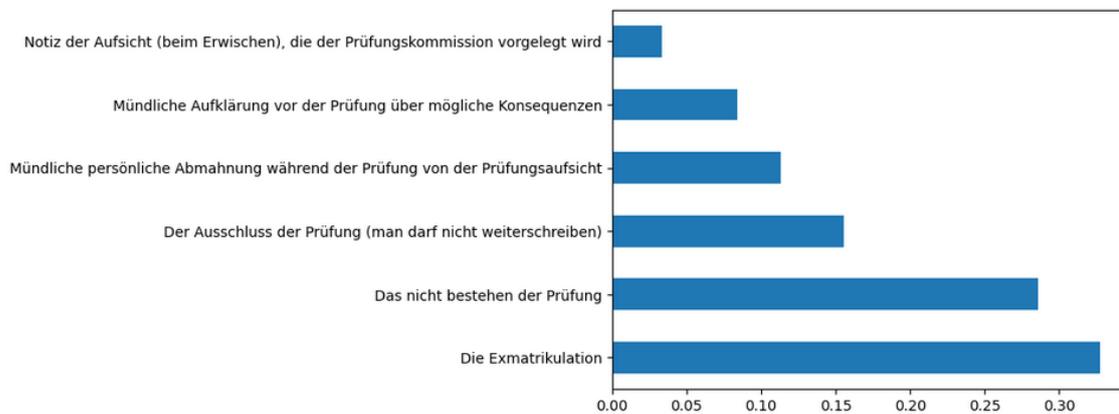


Abbildung 5-13 Antworten zur Frage, welche Konsequenzen Befragte vom Schummeln abhalten (Eigene Darstellung)

Eine weitere Frage der Umfrage beschäftigte sich mit Schummel-Methoden und wie bekannt sie unter den Teilnehmern sind. In der nachfolgenden Tabelle 5-2 ist dargestellt, wie viele Befragte die jeweilige Schummel-Methode kennen.

Schummel Methode	Anzahl der Befragten
Analoger Spickzettel	210
Kommunikation mit anderen Kommilitonen während der Prüfung	166
Einblendung von Inhalten auf einem anderen Gerät	121
Vorbereitete Texte werden abgelesen	119
Einblendung von Inhalten auf dem zweiten Bildschirm ohne Wissen des Prüfers	108
Übersetzungsprogramme	95
Lösungen von Dritten werden komplett kopiert	84
Einblendung von Inhalten auf den Hauptbildschirm ohne Wissen des Prüfers	83
Vortäuschung technischer Probleme	79
Manipulierte Prüfungsmaterialien	61
Prüfung wird von jemand anderen geschrieben	57
Audio Signal im Ohr	55
Virtuelle Kamera	31

Tabelle 5-2 Bekanntheit der Schummel Methoden (Eigene Darstellung)

Die bekanntesten Schummel Methoden sind der Analoge Spickzettel, die Kommunikation mit anderen während der Prüfung und das Einblenden von Inhalten auf anderen Geräten.

Der analoge Spickzettel würde von den Teilnehmern am meisten in den Take Home Exam's angewendet werden. So würden 104 Teilnehmer den Spickzettel in der Moodle Abgabe und 103 im Moodle Test als Täuschungsmittel benutzen. In der schriftlichen Zoom Prüfung würden 72 Teilnehmer diese Methode

verwenden, wohingegen 55 Teilnehmer die Methode in der mündlichen Zoom Prüfung anwenden würden. In einer Pool EDV Raum Prüfung würden 54 Teilnehmer der Umfrage, diese Methode als Täuschungsmittel nutzen (siehe Abbildung 5-14).

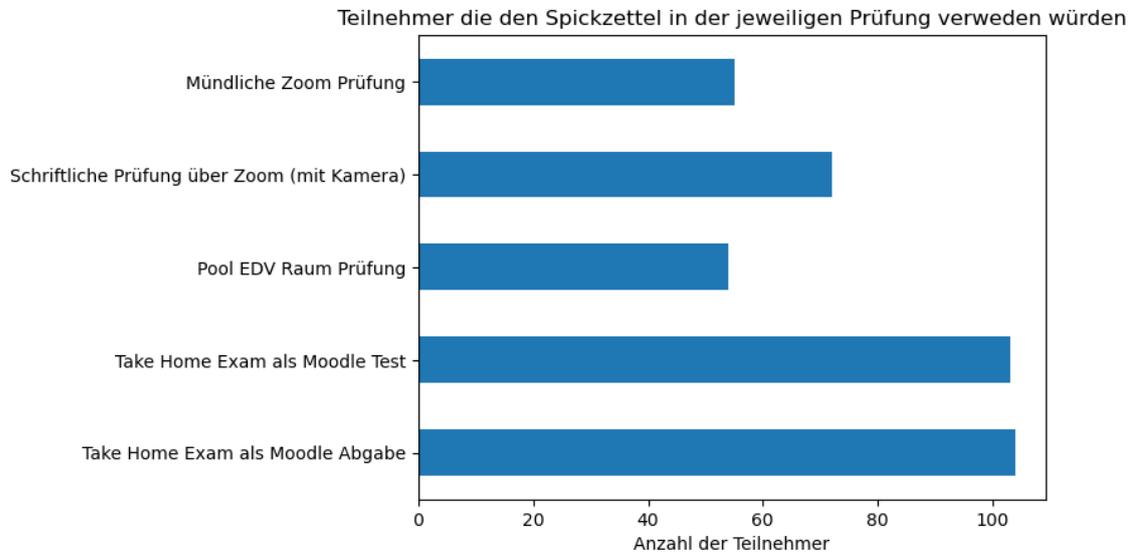


Abbildung 5-14 Verwendung des Spickzettels in verschiedenen Prüfungen (Eigene Darstellung)

Bei der Schummel Methode „Kommunikation mit anderen während der Prüfung“ würden ebenfalls die meisten Teilnehmer diese in den Take Home Exam’s verwenden (siehe Abbildung 5-15).

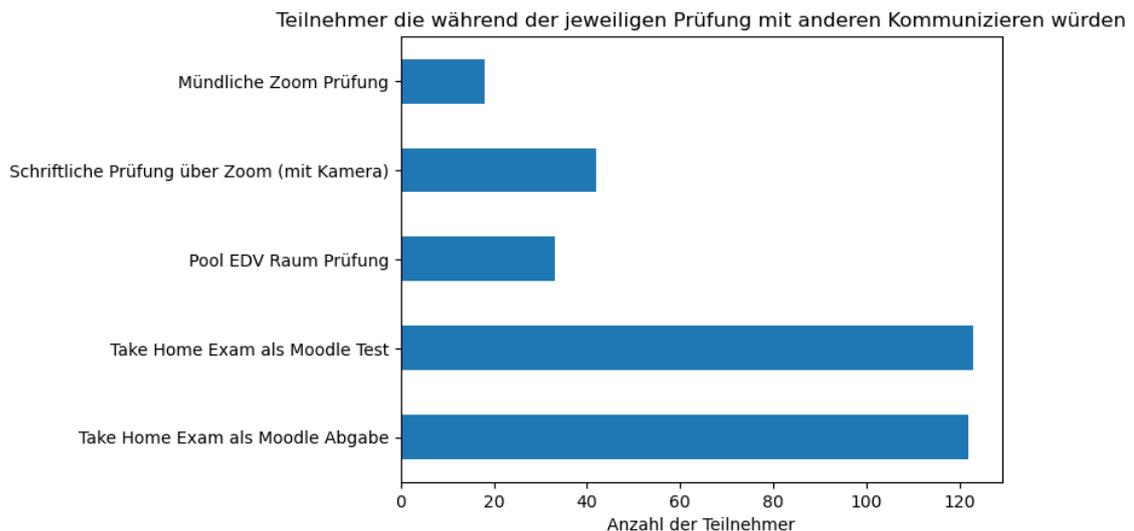


Abbildung 5-15 Verwendung der Kommunikation mit anderen während der Prüfung als Täuschungsmittel in verschiedenen Prüfungen (Eigene Darstellung)

Im Moodle Test wären 123 Personen bereit diese Methode zu verwenden, in der Moodle Abgabe 122 Teilnehmer der Umfrage. Während der schriftlichen Zoom Prüfung würden 42 Befragte, während der Pool EDV Raum Prüfung würden 33

Personen und bei einer mündlichen Zoom Prüfung würden 18 Personen zu diesem Täuschungsmittel zurückgreifen.

Aufgrund der begrenzten Länge der Arbeit, konnten nicht alle Argumente und Fragen in der deskriptiven Auswertung dargestellt werden. Es wurde eine Auswahl der aussagekräftigsten Argumente und Fragen dargestellt.

## 5.2. Clusterauswertung

Bei der Cluster-Auswertung wurden die Daten aus der deskriptiven Auswertung verwendet. Für das kMeans-Clustering wurden die demographischen Daten, der Studiengang, die privaten Aktivitäten, die Zufriedenheit und die Motivation, das tatsächliche Schummelverhalten und die Schummelhaltung verwendet. Nach Ausführung des Cluster-Algorithmus entstanden folgende drei Cluster (siehe Abbildung 5-16).

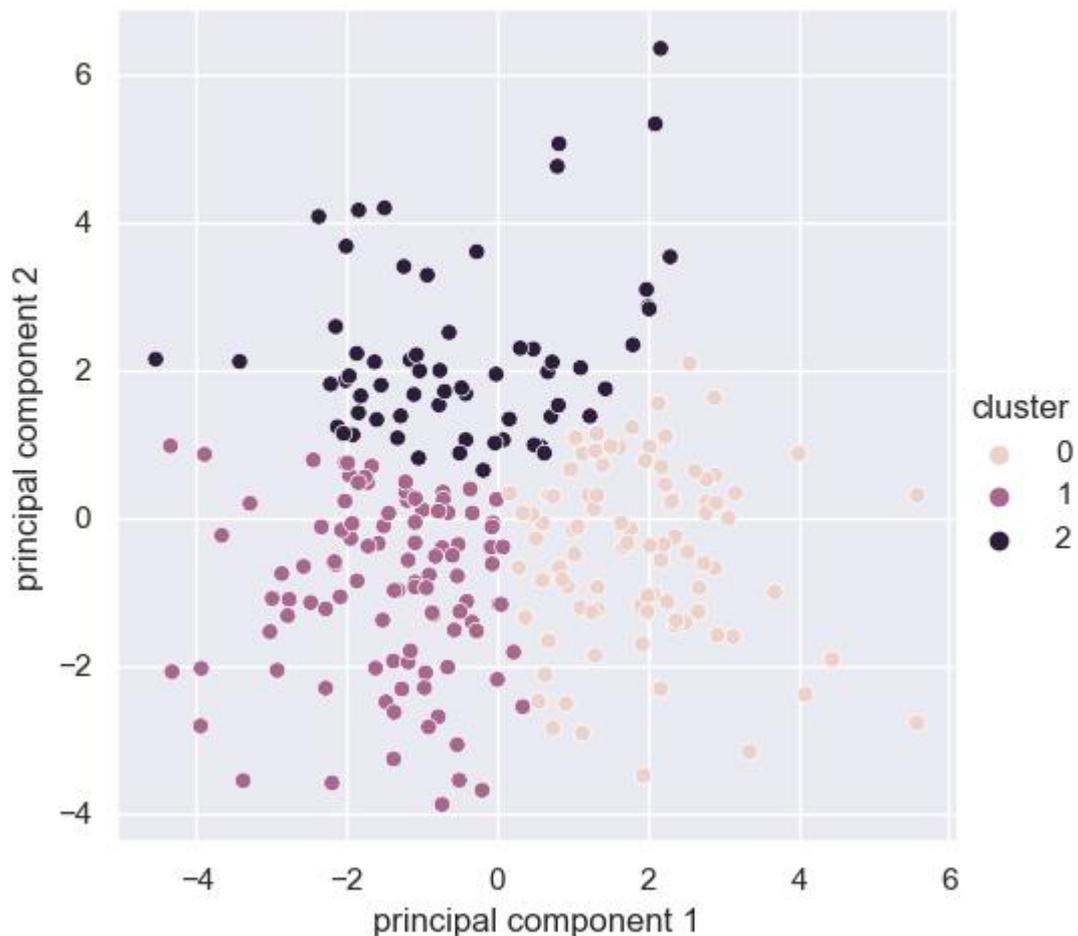


Abbildung 5-16 kMeans Clustering mit drei Clustern (Eigene Darstellung)

Die Auswertung der drei Cluster wurden mit dem Mittelwert durchgeführt. Insgesamt befinden sich 92 Teilnehmer der Umfrage in Cluster 0, 102 Teilnehmer in Cluster 1 und 63 Teilnehmer in Cluster 2.

Zuerst werden die Studiengänge ausgewertet, wie in Abbildung 5-17 dargestellt. Auf dem ersten Blick ist zu erkennen, dass im Cluster 0 alle Studiengänge enthalten sind, wohingegen bei Cluster 1 und 2 einige Studiengänge nicht enthalten sind. Erkennbar ist das im Cluster 0 die Studiengänge DSM, GPM, WIF und WING überdurchschnittlich häufig vertreten sind. Der Studiengang WIF befindet sich ausschließlich in diesem Cluster. Im Cluster 1 ist zu erkennen das BWL, DMP, IMUK und PHA häufig vertreten sind. Außerdem ist zu erkennen das in diesem Cluster die Studiengänge GMP, IMA und WIF nicht vertreten sind. Im dritten Cluster sind die Studiengänge DSM, DMP, IG, PHA, WIF nicht vertreten. Am häufigsten sind die Studiengänge BWG, DEM, IMA und WSP in Cluster 2 enthalten.

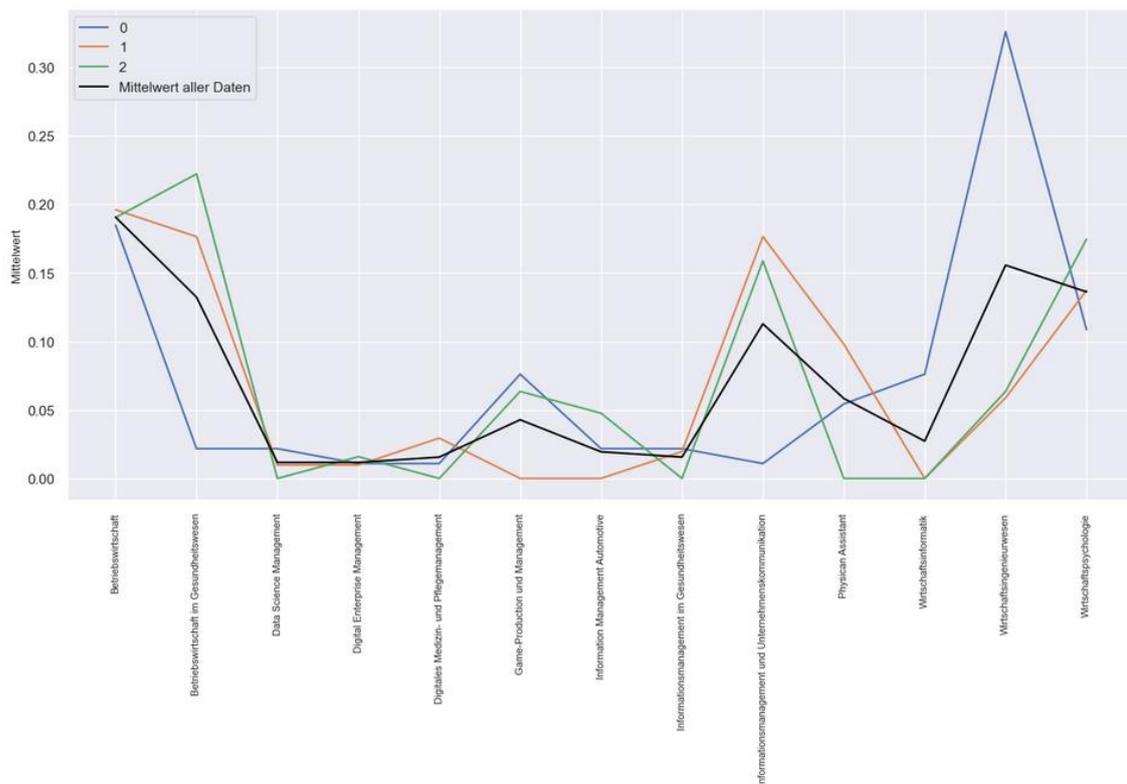


Abbildung 5-17 Studiengang-Auswertung der Cluster (Eigene Darstellung)

Die Cluster unterscheiden sich hinsichtlich des Notendurchschnittes, des Fachsemesters und des Alters nur minimal, wie in

	Cluster 0		Cluster 1		Cluster 2		alle Daten	
	M	SD	M	SD	M	SD	M	SD
<b>Notendurchschnitt</b>	2,15	0,48	2,06	0,48	2,25	0,46	2,14	0,48
<b>Fachsemester</b>	3,33	2,32	3,76	2,51	5,28	2,41	3,98	2,53
<b>Alter</b>	20,60	8,08	21,06	8,03	23,90	5,91	21,59	7,67

*Tabelle 5-3 Mittelwert und Standardabweichung des Notendurchschnitts, des Fachsemesters und des Alters aller Daten und der Clusterzugeordneten Teilnehmer (Eigene Darstellung)*

ersichtlich. Der Notendurchschnitt von Cluster 0 ist dem Mittelwert aller Daten am nächsten. Cluster 2 liegt über dem Mittelwert aller Daten und Cluster 1 liegt unter dem Mittelwert aller Daten.

Das Fachsemester von Cluster 2 liegt über dem Mittelwert aller Daten, wohingegen Cluster 0 unter dem Mittelwert aller Daten liegt. Cluster 1 liegt am nächsten am Mittelwert aller Daten, allerdings auch unter diesem. Das mittlere Alter liegt in Cluster 1 fast auf dem Mittelwert aller Daten. Das Alter in Cluster 0 liegt unterhalb und das Alter in Cluster 2 liegt über dem Mittelwert aller Daten.

	Cluster 0		Cluster 1		Cluster 2		alle Daten	
	M	SD	M	SD	M	SD	M	SD
<b>Notendurchschnitt</b>	2,15	0,48	2,06	0,48	2,25	0,46	2,14	0,48
<b>Fachsemester</b>	3,33	2,32	3,76	2,51	5,28	2,41	3,98	2,53
<b>Alter</b>	20,60	8,08	21,06	8,03	23,90	5,91	21,59	7,67

*Tabelle 5-3 Mittelwert und Standardabweichung des Notendurchschnitts, des Fachsemesters und des Alters aller Daten und der Clusterzugeordneten Teilnehmer (Eigene Darstellung)*

Hinsichtlich der Untersuchung der Cluster auf ihre Geschlechtsverteilung ist ersichtlich, dass in Cluster 1 der weibliche Anteil sehr hoch ist. Im Gegensatz dazu ist der männliche Anteil in Cluster 0 am höchsten. Im Cluster 2 sind die Anteile der männlichen und weiblichen Teilnehmer vergleichbar (siehe Abbildung 5-18), wobei erkennbar ist dass der männliche Wert unter dem Mittelwert aller Daten liegt und der weibliche Wert über den Mittelwert aller liegt.

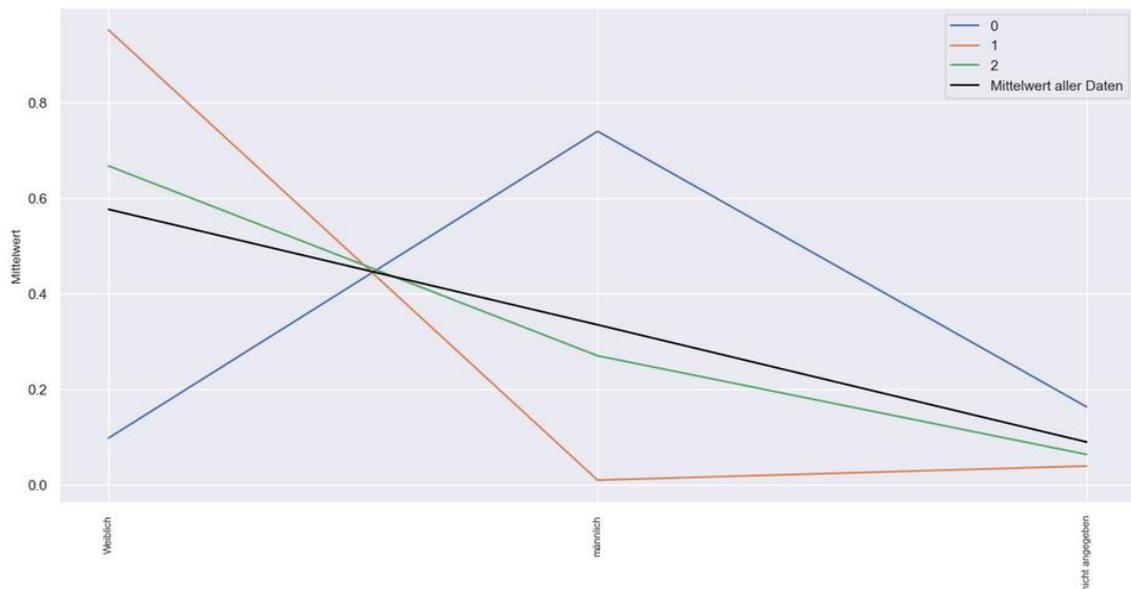


Abbildung 5-18 Geschlechtsverteilung der Cluster (Eigene Darstellung)

Im Fall der Untersuchung der Wohnsitze ist erkennbar, dass Cluster 0 signifikant ähnlich zu den Mittelwerten aller Wohnorte ist. Im Cluster 1 hingegen ist ersichtlich, dass überdurchschnittlich viele Teilnehmer bei den Eltern und im Studentenwohnheim wohnen. In einer eigenen Wohnung oder einem WG-Zimmer wohnten hingegen unterdurchschnittlich viele. Cluster 2 ist invers zu Cluster 1. Eine eigene Wohnung und ein WG-Zimmer haben in diesem Cluster häufig viele, bei den Eltern und im Studentenwohnheim wohnen unterdurchschnittlich wenige (siehe Abbildung 5-19).

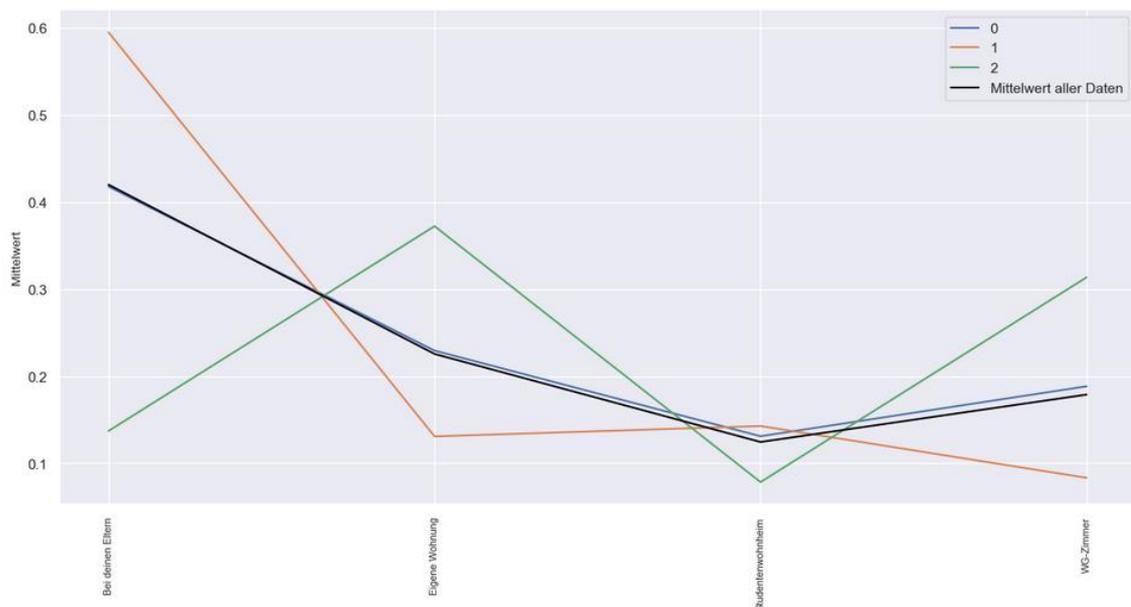


Abbildung 5-19 Wohnorte in den Clustern (Eigene Darstellung)

In der unteren Graphik (Abbildung 5-20) ist ersichtlich, dass Cluster 0 und Cluster 1 bezüglich der Verteilung der Zufriedenheit mit dem Studium ähnlich aufgeteilt sind und weisen einen ähnlichen Trend wie der Mittelwert aller Daten auf. Sie haben die höchsten Mittelwerte bei „Stimme zu“.

Hingegen ist die Verteilung bei Cluster 2 stark vertreten bei „Stimme nicht zu“ und „Weder noch“. Auch dieser Cluster folgt dem Trend der Mittelwerte aller Daten. Die Antworten „Stimme voll und ganz zu“, „Stimme überhaupt nicht zu“ und „Stimme zu“ sind unterdurchschnittlich vertreten.

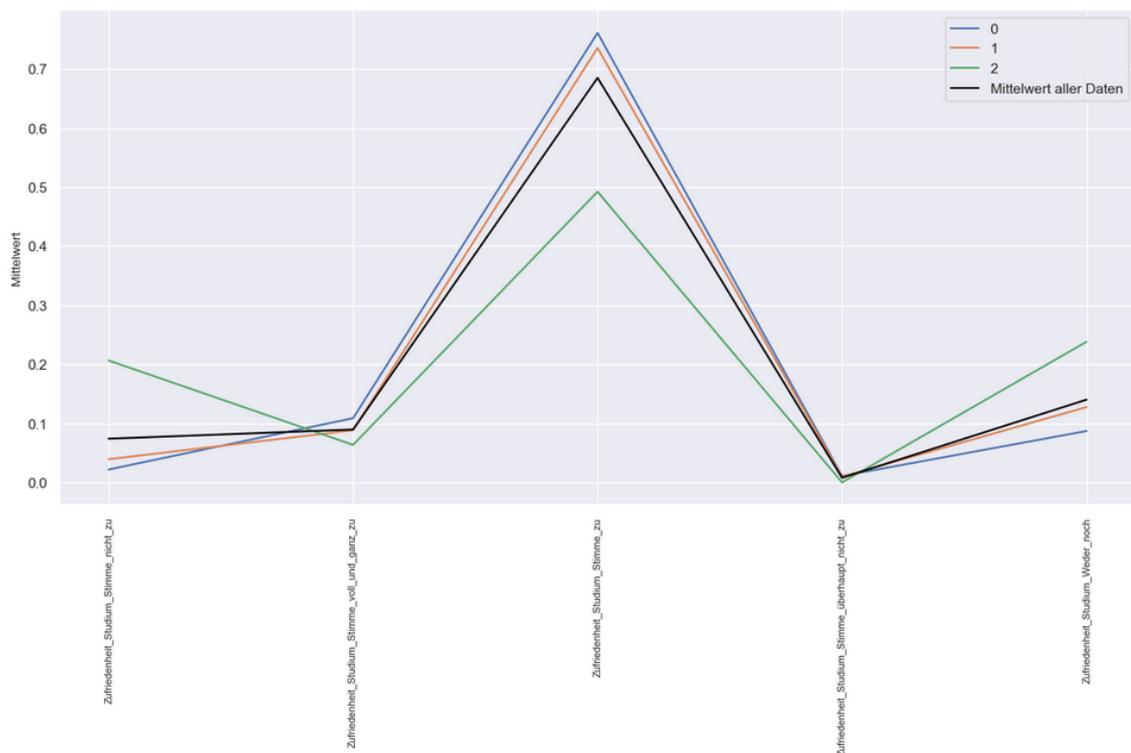


Abbildung 5-20 Clusterauswertung der Zufriedenheit mit dem Studium (Eigene Darstellung)

Ähnlich sieht die Verteilung bei der Auswertung der Zufriedenheit mit der Hochschule aus. Hier weisen Cluster 0 und Cluster 1 ebenfalls einen ähnlichen Trend wie der Mittelwert aller Daten aus. Die höchsten Mittelwerte liegen bei der Antwort „Stimme zu“. Unterschiede zwischen Cluster 0 und Cluster 1 befinden sich bei den antworten „Stimme voll und ganz zu“ und „weder noch“. Nur die Antwort „weder noch“ ist im Cluster 0 auch überdurchschnittlich oft vertreten, alle anderen Antworten sind unterdurchschnittlich oft vorhanden. Bei Cluster 1 ist nur die Antwort „Stimme nicht zu“ auch überdurchschnittlich vertreten, alle anderen liegen unter dem Mittelwert aller Daten. Der Cluster 2 weicht vom Trend ab. In diesem Cluster sind die Antworten „Stimme nicht zu“, „Stimme überhaupt nicht zu“ und „weder noch“ überdurchschnittlich vertreten. Die Antworten „Stimme voll

und ganz zu“ und „Stimme zu“ sind unterdurchschnittlich vertreten, wie in Abbildung 5-21 zu erkennen ist.

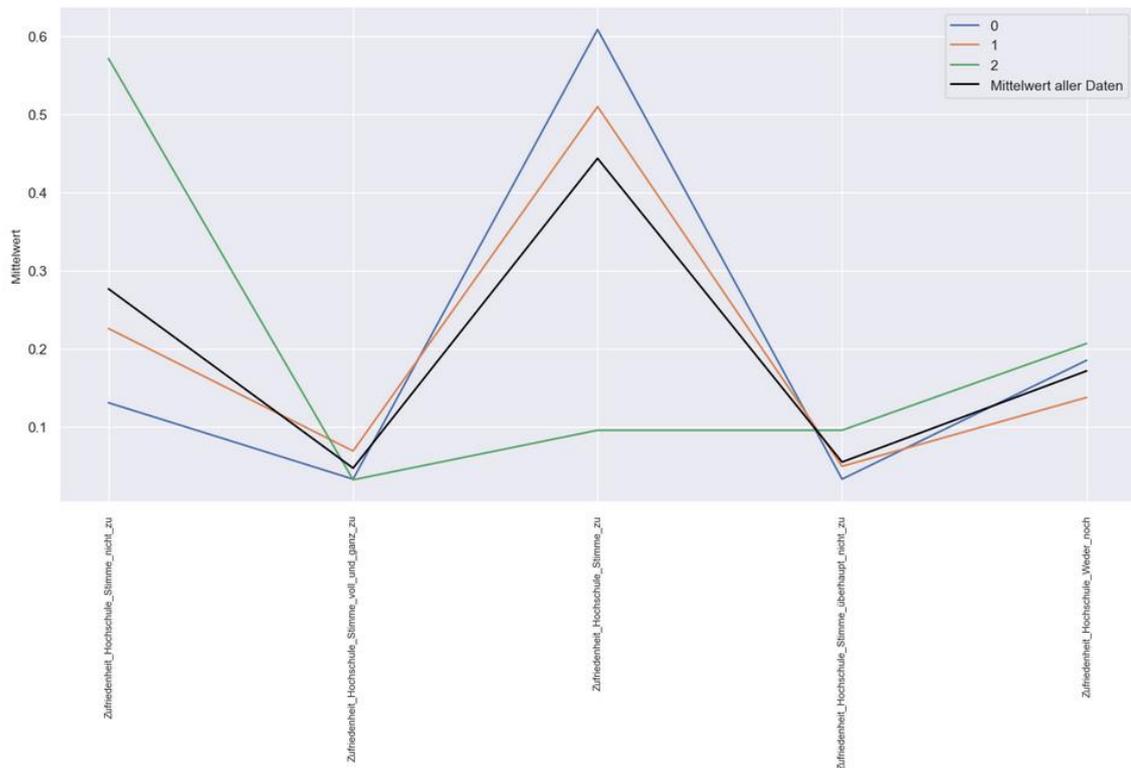


Abbildung 5-21 Clusterauswertung der Zufriedenheit mit der Hochschule (Eigene Darstellung)

Die Cluster unterscheiden sich hinsichtlich der Motivation nur ein wenig, wie in Abbildung 5-22 dargestellt. Wie in den vorherigen Auswertungen liegen die Mittelwerte der Cluster 0 und Cluster 1 sehr nahe an den Mittelwerten aller Daten. Hierbei ist „motiviert“ die Antwort mit den höchsten Mittelwerten. In Cluster 2 sind die überdurchschnittlichen Mittelwerte bei den Antworten „sehr unmotiviert“, „unmotiviert“ und „weder noch“ zu finden.

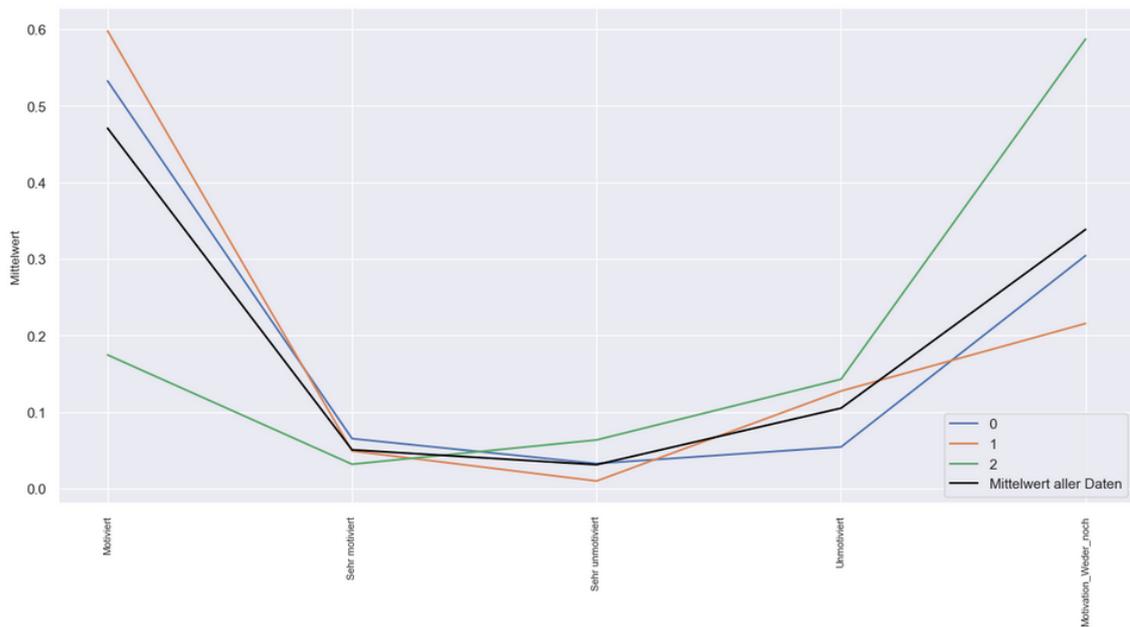


Abbildung 5-22 Clusterauswertung der Motivation (Eigene Darstellung)

In Abbildung 5-23 ist die Verteilung des Leistungsdrucks auf die Cluster sichtbar. Hier ist zu erkennen, dass Cluster 0 den höchsten Mittelwert bei den Antworten „Stimme zu“ und den niedrigsten bei den Antworten „Stimme überhaupt nicht zu“ besitzt. Zusätzlich enthält dieser Cluster überdurchschnittlich „Stimme voll und ganz zu“ Antworten, jedoch besitzt Cluster 2 in diesem Punkt den höchsten Mittelwert. Cluster 1 besitzt den höchsten Mittelwert bei den „Stimme nicht zu“ und „weder noch“ Antworten. Unterdurchschnittlich viele Antworten enthält dieser Cluster bei den „Stimme voll und ganz zu“ Antworten. Im Cluster 2 befinden sich überdurchschnittlich viele „Stimme überhaupt nicht zu“ Antworten, zudem sind auch überdurchschnittlich viele „weder noch“ Stimmen in diesem zu finden.

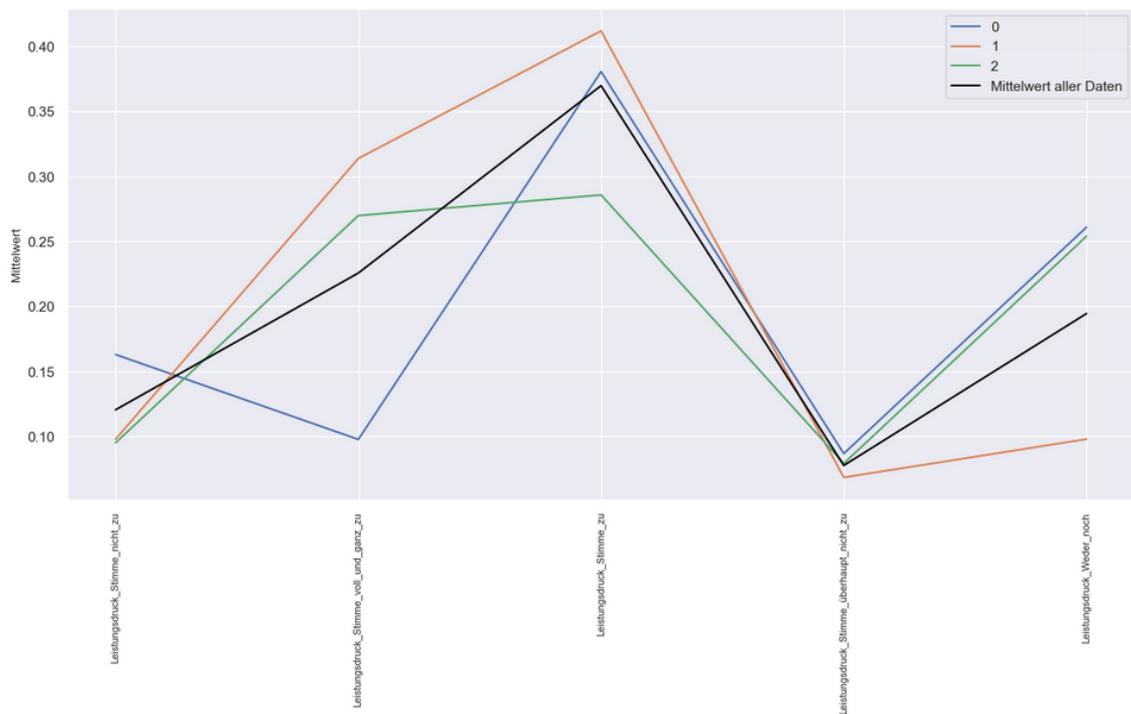


Abbildung 5-23 Clusterauswertung des Leistungsdrucks (Eigene Darstellung)

Die Verteilung der Freizeitaktivitäten ist in Abbildung 5-24 abgebildet. Zu sehen ist hierbei das Cluster 0 die Aktivitäten Fußball, Basketball, Tennis, Klettern, Schach und Poker überdurchschnittlich oft enthält. Auch zu erkennen ist das alle Freizeitaktivitäten vertreten sind. Cluster 1 besitzt keine Antworten bei Basketball und Poker. In diesem Cluster sind die überdurchschnittlichen Mittelwerte bei Tanzen, Yoga, Fitnessstudio, Pilates, Handball, Joggen/Laufen, Freunde treffen, Lesen, Malen, Kino und Bar/Club zu finden. Dem Cluster 2 ist die Freizeitaktivität Poker nicht zugeordnet. In diesem Cluster sind Schwimmen und Reiten stark vertreten.

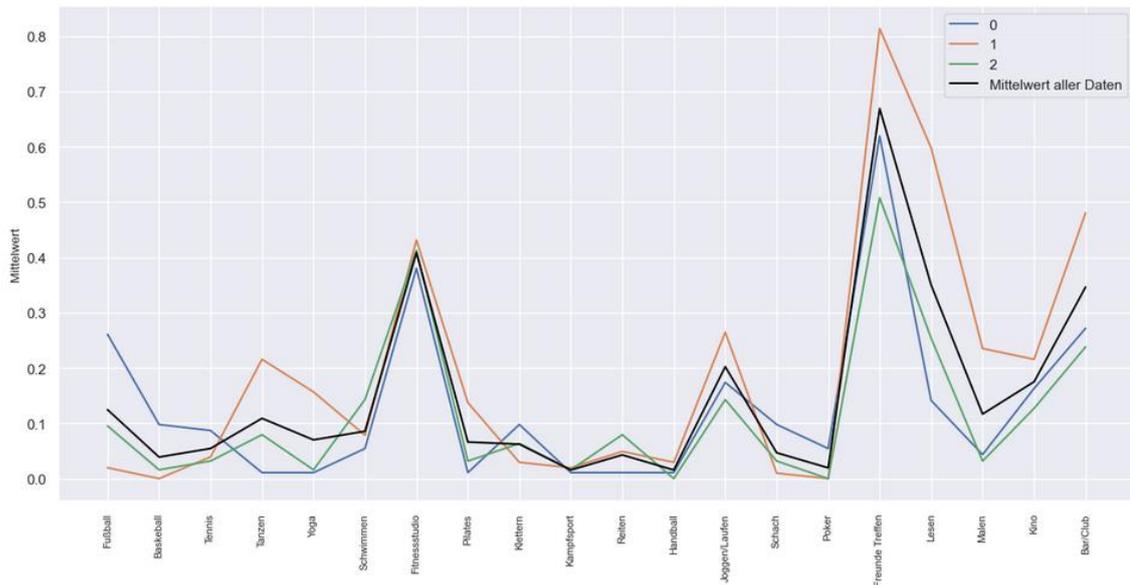


Abbildung 5-24 Clusterauswertung der Freizeitaktivitäten (Eigene Darstellung)

Bei der Auswertung des Schummelverhalten und der Schummelhaltung ist zu sehen, dass Cluster 0 und Cluster 1 dem Trend des Mittelwerts aller Daten folgen. Auch erkennbar ist, dass in jedem Cluster geschummelt wurde, allerdings wurde im Cluster 2 besonders häufig geschummelt.

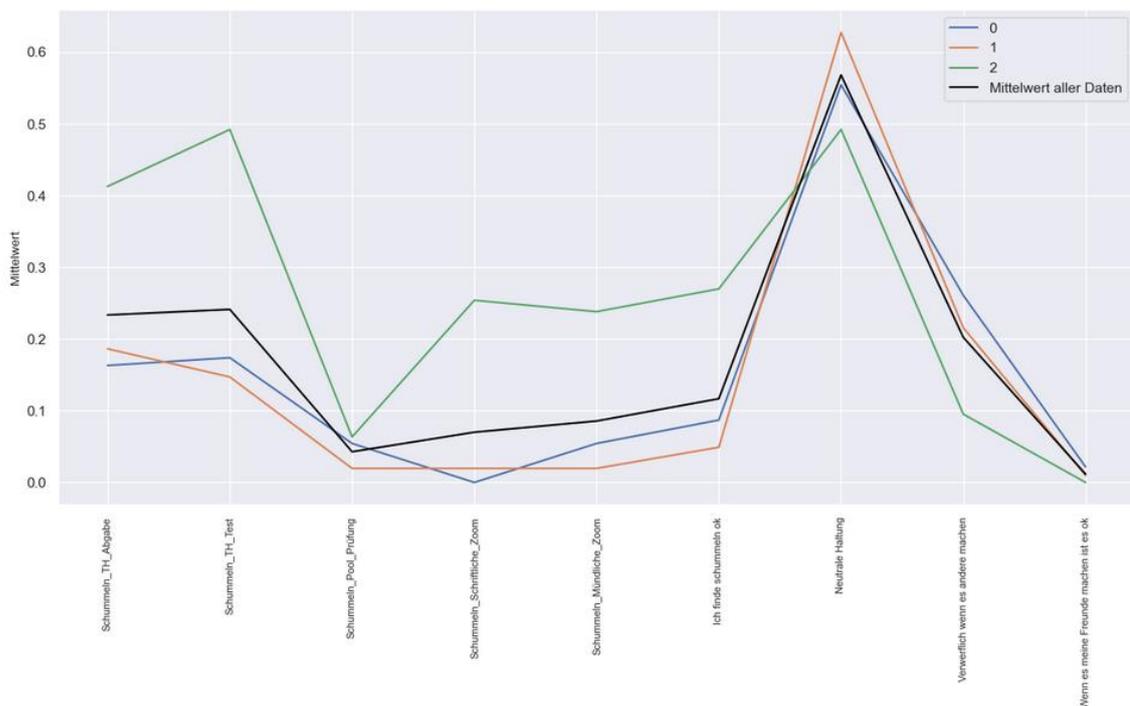


Abbildung 5-25 Clusterauswertung der Schummel Haltung und des tatsächlichen Schummel Verhalten (Eigene Darstellung)

Es kann festgehalten werden, dass Cluster 2 bei allen Prüfungen überdurchschnittlich oft schummelt. Zudem ist die Schummelhaltung überdurchschnittlich positiv. Bei der Auswertung von Cluster 1 ist zu sehen, dass

sie in allen Prüfungen unterdurchschnitt oft schummeln. Die Schummelhaltung von Cluster 1 ist eher neutral und verwerflich.

Cluster 0 schummelt nur in der EDV Raum Pool Prüfung überdurchschnittlich oft. In allen anderen abgefragten Prüfungen wird wie bei Cluster 1 unterdurchschnittlich oft geschummelt. Die Schummelhaltung von Cluster 0 ist eher ablehnend, aber es ist für sie ok, wenn Freunde schummeln.

## 6. Fazit und Empfehlung

Abschließend kann festgehalten werden, dass die Ergebnisse der Befragung eindeutig der Fakultät WW zugeordnet werden können. Die meisten Teilnehmerinnen und Teilnehmer studieren BWL. Das Alter variierte zwischen 18 und 56 Jahren, die meisten Studierenden waren zwischen 22 und 23 Jahren alt. Der durchschnittliche Notendurchschnitt lag bei 2,14 und mehr als die Hälfte der Befragten waren weiblich. Die meisten Studierenden sind mit ihrem Studium zufrieden und haben eine positive Einstellung zur Hochschule. Der Leistungsdruck wird von vielen Studierenden als hoch empfunden, jedoch sind sie auch hoch motiviert.

In Bezug auf Freizeitaktivitäten zeigte sich, dass die meisten Teilnehmerinnen und Teilnehmer Freunde treffen, ins Fitnessstudio gehen, Lesen oder in Bars und Clubs ausgehen. Die meisten Studierenden besuchen fast alle oder alle Vorlesungen. Im Bereich der Prüfungen wurden verschiedene Prüfungsformen und die Haltung der Teilnehmerinnen und Teilnehmer zum Schummeln untersucht. Dabei kam heraus, dass die meisten Teilnehmerinnen und Teilnehmer Präsenz Prüfungen bevorzugen und dass die Konsequenzen des Schummelns den meisten bekannt sind. Über die Hälfte der Teilnehmer haben eine neutrale Haltung zum Schummeln. Die bevorzugten Prüfungen, in denen geschummelt wurde, sind Take-Home-Exams. Die beliebtesten Schummel Methoden sind der analoge Spickzettel und die Kommunikation mit anderen während der Prüfung.

Eine Clusteranalyse mithilfe des kMeans-Algorithmus wurde durchgeführt, um Muster in den Daten zu identifizieren. Daraus resultierten drei Cluster mit

unterschiedlichen Merkmalen. Eine Auswertung der Cluster ergab Unterschiede hinsichtlich der Studiengänge, des Notendurchschnitts, des Fachsemesters, des Alters, der Geschlechtsverteilung und der Wohnsituation. Zusätzlich wurde die Zufriedenheit mit der Hochschule, dem Studium, die Motivation, der Leistungsdruck, die Freizeitaktivität, das Schummelverhalten und die Haltung zum Schummeln hinsichtlich der drei Cluster untersucht. Eine Zusammenfassung der Ergebnisse findet sich in Tabelle 6-1.

	<b>Cluster 0</b>	<b>Cluster 1</b>	<b>Cluster 2</b>
<b>Studiengang</b>	DSM, GMP, IG, WIF, WING	BWL, DMP, IMUK, PHA	BGW, DEM, IMA, WSP
<b>Geschlecht</b>	männlich	weiblich	Beides ca. gleich verteilt
<b>Wohnort</b>	Bei den Eltern, Studentenwohnheim	Bei den Eltern, WG-Zimmer	Eigene Wohnung, WG-Zimmer
<b>Zufriedenheit Studium</b>	Stimme zu		Stimme nicht zu, Weder noch
<b>Zufriedenheit Hochschule</b>	Stimme zu	Stimme voll und ganz zu	Stimme nicht zu, Stimme überhaupt nicht zu, weder noch
<b>Motivation</b>	Sehr motiviert	Motiviert	sehr unmotiviert, unmotiviert, weder noch
<b>Leistungsdruck</b>	Stimme nicht zu, Stimme überhaupt nicht zu, weder noch	Stimme voll und ganz zu, Stimme zu	Stimme voll und ganz zu, weder noch
<b>Hobbies</b>	Fußball, Basketball, Tennis, Klettern, Schach, Poker	Tanzen, Yoga, Fitnessstudio, Pilates, Kampfsport, Handball, Joggen/Laufen, Freunde treffen, Lesen, Malen, Kino, Bar/Club	Schwimmen, Reiten

*Tabelle 6-1 Zusammenfassung der Clusterergebnisse (Eigene Darstellung)*

Nach der Auswertung lässt sich festhalten, dass Cluster 2 am häufigsten schummelt und eine positive Einstellung zum Schummeln hat. Cluster 1 hingegen schummelt eher selten und hat eine neutrale und ablehnende Einstellung zum Schummeln. Cluster 0 schummelt nur in einer bestimmten Prüfung überdurchschnittlich oft und hat eine ablehnende Einstellung zum Schummeln, aber akzeptiert es, wenn Freunde schummeln.

Die Ergebnisse der vorliegenden Arbeit sind von hoher Bedeutung für die Wissenschaft und die Praxis, da sie einen wertvollen Einblick in die Zufriedenheit, Motivation, Leistungsdruck, Freizeitaktivitäten und weitere Einflussfaktoren von Studierenden an der Hochschule Neu-Ulm liefern. Die Studie zeigt, dass es Unterschiede zwischen den verschiedenen Clustern von Studierenden gibt, die sich in ihren Einstellungen und Aktivitäten unterscheiden.

Es sollte beachtet werden, dass die Studie nur an einer einzigen Hochschule durchgeführt wurde und eine große Anzahl von Teilnehmern aus der Fakultät für Wirtschaftswissenschaften stammte, wodurch die Ergebnisse möglicherweise nicht für die gesamte Hochschule repräsentativ sind. Wenn diese Art von Forschung an mehreren Hochschulen durchgeführt wird, können Vergleiche zwischen den verschiedenen Institutionen gezogen werden, um Gemeinsamkeiten und Unterschiede aufzuzeigen.

Eine Schwäche der vorliegenden Arbeit besteht darin, dass die Umfrage auf freiwilliger Basis durchgeführt wurde, wodurch einige Lücken in den Daten entstanden sind. Zur Behebung dieser Lücken wurden Mittelwerte und Nullen verwendet, was zu einer möglichen Verfälschung der Ergebnisse führen könnte. Daher empfiehlt es sich, in zukünftigen Studien bestimmte Fragen als obligatorische Angaben zu kennzeichnen, um das Auftreten von Lücken zu minimieren.

Basierend auf den Ergebnissen einer neuen Umfrage mit den Professoren der Hochschule Neu-Ulm ergaben sich neue Erkenntnisse zur Aktualität des Themas. Insbesondere ergab die Umfrage, dass die Hälfte der teilnehmenden Professoren auch weiterhin Online Prüfungen anbieten werden, wobei am häufigsten Take-Home Moodle Abgaben und mündliche Prüfungen zum Einsatz kommen. Zur Vermeidung von Schummeln setzen viele der Professoren auf Verständnisfragen oder Open-Book Prüfungen.

Darüber hinaus können die Ergebnisse dieser Arbeit von praktischem Nutzen für die Hochschulverwaltung und -planung sein. Indem sie das Verhalten und die Bedürfnisse der Studierenden besser verstehen, können Hochschulen Maßnahmen ergreifen, um das Studium und die damit verbundenen Erfahrungen zu verbessern. Zu diesen Maßnahmen zählt auch die Entwicklung von Freizeitmöglichkeiten und die Verminderung des Leistungsdrucks. Darüber

hinaus sollte die Prüfungsform geändert werden, um Betrugsversuche zu vermeiden und um das Schummeln bei Online Prüfungen obsolet zu machen.

Eine Änderung der Prüfungsform könnte durch alternative Prüfungsformate entstehen, die das Wissen und die Fähigkeiten der Studierenden während des Semesters kontinuierlich bewerten. Solche alternativen Prüfungsformate könnten beispielsweise Portfolio-Prüfungen, reine Verständnisfragen oder Gruppenarbeiten sein, die darauf abzielen, die Anwendung und Reflexion des Wissens zu fördern und das Auswendiglernen von Informationen zu reduzieren. Durch die Implementierung solcher Prüfungsformate kann der Leistungsdruck reduziert und das Schummeln in Online Prüfungen minimiert werden.

Als weitere Empfehlung ist es wichtig, dass Studierende über den richtigen Umgang mit KI-Tools wie ChatGPT<sup>35</sup> und anderen Technologien aufgeklärt werden, um ungewolltes Schummelverhalten zu vermeiden. Der Fokus sollte dabei nicht darauf liegen, alle KI-Tools zu verbieten, sondern vielmehr darauf, ein Bewusstsein für ihre Verwendung zu schaffen und den Studierenden den verantwortungsvollen Umgang damit beizubringen.

Zusammenfassend kann gesagt werden, dass die Ergebnisse dieser Arbeit wertvolle Erkenntnisse liefern, die für die Wissenschaft und die Praxis von Nutzen sein können. Eine Erweiterung auf mehrere Hochschulen kann dazu beitragen, ein umfassenderes Verständnis der Einflussfaktoren der Studierenden zu erlangen und Maßnahmen zu entwickeln, um das Studium zu verbessern.

---

<sup>35</sup> ChatGPT ist ein textbasiertes Dialogsystem, in Form eines Chatbots, der auf maschinelles Lernen basiert.

## Literaturverzeichnis

Alpaydin, Ethem (2022): Maschinelles Lernen. 3., aktualisierte und erweiterte Auflage. Berlin, Boston: De Gruyter Oldenbourg (De Gruyter Studium). Online verfügbar unter <https://www.degruyter.com/isbn/9783110740141>.

Bendel, Oliver (2019): Maschine Learning. Definition: Was ist "Maschine Learning"? Hg. v. Springer Fachmedien Wiesbaden GmbH. Gabler Wirtschaftslexikon. Online verfügbar unter <https://wirtschaftslexikon.gabler.de/definition/machine-learning-120982/version-370915>, zuletzt geprüft am 31.10.2022.

Breitenbach, Andrea (2021): Digitale Lehre in Zeiten von Covid-19: Risiken und Chancen. Hg. v. DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation. Marburg.

Ertel, Wolfgang (2021): Grundkurs Künstliche Intelligenz. Eine praxisorientierte Einführung. 5. Auflage. Wiesbaden, Heidelberg: Springer Vieweg (Lehrbuch). Online verfügbar unter <http://www.springer.com/>.

Goertz, Lutz; Hense, Julia (2021): Studie zur Veränderungsprozessen in Unterstützungsstrukturen für Lehre an deutschen Hochschulen in der Corona-Krise. Arbeitspapier Nr. 56. Berlin: Hochschulforum Digitalisierung. Unter Mitarbeit von Hendrik Metz und Michael Georgi. Hg. v. Geschäftsstelle Hochschulforum Digitalisierung beim Stifterverband für die Deutsche Wissenschaft e.V. [hochschulforumdigitalisierung.de/publikationen](https://hochschulforumdigitalisierung.de/publikationen). Online verfügbar unter [hochschulforumdigitalisierung.de/publikationen](https://hochschulforumdigitalisierung.de/publikationen).

Hirschle, Jochen (2021): Machine Learning für Zeitreihen. Einstieg in Regressions-, ARIMA- und Deep Learning-Verfahren mit Python. München: Hanser (Hanser eLibrary). Online verfügbar unter <https://www.hanser-elibrary.com/doi/book/10.3139/9783446468146>.

Janke, Stefan; Rudert, Selma C.; Petersen, Änne; Fritz, Tanja M.; Daumiller, Marin (2021): Cheating in the wake of COVID-19: How dangerous is ad-hoc online testing for academic integrity? In: *Computers and Education Open* 2, S. 100055. DOI: 10.1016/j.caeo.2021.100055.

Kormann, Julia; Eßbach, Michaela; Köppl, Sonja (2022): ii.Oo Digitales kompetenz-orientiertes Prüfen. Hg. v. Hochschule für angewandte Wissenschaften Neu-Ulm.

Marco Arnold (o.J.): Machine Learning - Das maschinelle Lernen. Hg. v. ARNOLD IT Systems GmbH & Co. KG. <https://www.arnold-it.com/>. Online verfügbar unter <https://www.arnold-it.com/digitalisierung/machine-learning-maschinelles-lernen/>, zuletzt geprüft am 11.11.2022.

Ng, Annalyn; Soo, Kenneth (2018): Data Science - was ist das eigentlich?! Algorithmen des maschinellen Lernens verständlich erklärt. Berlin, Heidelberg: Springer.

Nguyen, Chi Nhan; Zeigermann, Oliver (2021): Machine Learning - kurz & gut. 2. Auflage. Heidelberg: O'Reilly (O'Reillys Taschenbibliothek).

o.V. (o.J.): Lehrbuch Psychologie. Likert-Skala. Hg. v. Springer Fachmedien Wiesbaden. Online verfügbar unter <https://lehrbuch-psychologie.springer.com/glossar/likert-skala-1>, zuletzt geprüft am 04.02.2023.

Papp, Stefan; Weidinger, Wolfgang; Munro, Katherine; Ortner, Bernhard; Cadonna, Annalisa; Langs, Georg et al. (2022): Handbuch Data Science und KI. Mit Machine Learning und Datenanalyse Wert aus Daten generieren. 2. Auflage. München: Hanser (Hanser eLibrary). Online verfügbar unter <https://www.hanser-elibrary.com/doi/book/10.3139/9783446472457>.

Patel, Ankur A. (2020): Praxisbuch Unsupervised Learning. Machine-Learning-Anwendungen für ungelabelte Daten mit Python programmieren. 1. Auflage, deutsche Ausgabe. Heidelberg: O'Reilly.

Wuttke, Laurenz (o. J.): Was ist Deep Learning? datasolut GmbH. [www.datasolut.com](http://www.datasolut.com). Online verfügbar unter <https://datasolut.com/was-ist-deep-learning/#Unterschied-Machine-Learning-und-Deep-Learning>, zuletzt geprüft am 31.10.2022.

## Anhang

1. Umfrage.....	X
2. Auszug aus der bearbeiteten Exceldatei mit den Umfrageergebnissen...	XXI
3. Jupiter Notebook der deskriptiven Auswertung .....	XXII
4. Auszug aus der Exceldatei für das Clustering.....	XXXV
5. Clusterverfahren kMeans Code .....	XXXVI

# 1. Umfrage

Liebe Mitstudenten,

im Rahmen meiner Bachelorarbeit führe ich eine Umfrage zur Untersuchung von Einflussfaktoren auf die Studienleistung bei online Prüfungen durch.

Ich wäre euch sehr dankbar, wenn Ihr an meiner Umfrage teilnehmen würdet! Alle Antworten werden anonym behandelt und es können keine Rückschlüsse auf euch gezogen werden.

Die Umfrage dauert ca. 10min.

Vielen Dank für eure Unterstützung!

Stefanie

## Teil A: Allgemeine Fragen zum Studium

A1. Welchen Studiengang besuchst du?

- Betriebswirtschaft
- Betriebswirtschaft im Gesundheitswesen
- Digital Enterprise Management
- Digitales Medizin- und Pflegemanagement
- Data Science Management
- Game-Production und Management
- Information Management Automotive
- Informationsmanagement im Gesundheitswesen
- Physican Assistant
- Systems Engereering
- Wirtschaftsinformatik
- Wirtschaftsingenieurwesen
- Wirtschaftspsychologie
- Sonstiges

Sonstiges

**A2. In welchem Fachsemester befindest du dich?**

- 1. Fachsemester
- 2. Fachsemester
- 3. Fachsemester
- 4. Fachsemester
- 5. Fachsemester
- 6. Fachsemester
- 7. Fachsemester
- 8. Fachsemester
- 9. Fachsemester
- 10. Fachsemester und höher

**A3. Welchen Notendurchschnitt hast du aktuell?**

- 1,0-1,5
- 1,6-2,0
- 2,1-2,5
- 2,6-3,0
- 3,1-3,5
- 3,6-4,0

**Teil B: Allgemeine Fragen zum Privatleben**

**B1. Bist du mit deinem Studium bisher zufrieden?**

- Stimme voll und ganz zu
- Stimme zu
- Weder noch
- Stimme nicht zu
- Stimme überhaupt nicht zu

**B2. Bist du im Allgemeinen mit der Organisation der Hochschule zufrieden?**

- Stimme voll und ganz zu
- Stimme zu
- Weder noch
- Stimme nicht zu
- Stimme überhaupt nicht zu

**B3. Fühlst du dich in deinem Studium motiviert, z.B. durch Dozenten oder Mitstudierende?**

- Sehr motiviert
- Motiviert
- Weder noch
- Unmotiviert
- Sehr unmotiviert

**B4. Wie viele Stunden verbringst du in der Woche mit der Vor- und Nachbearbeitung deines Studiums?**

- 0-5
- 6-10
- 11-15
- 16-20
- 21-25
- 26-30
- 31-35
- 36-40

**B5. Wie regelmäßig besuchst du die Vorlesung im Semester?**

- Ich besuche alle Vorlesungen
- Ich besuche fast alle Vorlesungen
- Ich lasse regelmäßig Vorlesungen ausfallen
- Ich gehe nicht in Vorlesungen

**B6. Verspürst du einen Leistungsdruck, z.B. sehr gute Noten um nach dem Abschluss einen guten Job zu finden?**

Stimme voll und ganz zu

Stimme zu

Weder noch

Stimme nicht zu

Stimme überhaupt nicht zu

---

**B7. Wie verbringst du deine Freizeit? (Sport/Hobby/...)**

Fußball

Basketball

Tennis

Tanzen

Yoga

Schwimmen

Fitnessstudio

Pilates

Klettern

Kampfsport

Reiten

Handball

Joggen/Laufen

Schach

Poker

Freunde Treffen

Lesen

Malen

Kino

Bar/Club

Sonstiges

Sonstiges

---

**B8. Wie viel Zeit nimmt dein Hobby/ deine Sportart in der Woche in Anspruch?**

0-3

4-7

8-11

12-15

16-19

20-23

24-27

28-31

32-35

36-39

40 und mehr

---

**B9. Welche Interessensgebiete hast du?**

Politik

Basteln

Handwerk

Kunst

Mode

Musik

Sonstiges

Sonstiges

**B10. Welchen Nebenjob hast du? (studentische Tätigkeit, Verkauf, Handwerk,...)**

**B11. Hat dein Nebenjob etwas mit deinem Studium zu tun?**

Ja

Nein

**B12. Falls du neben deinem Studium einen Job ausübst - Wie viele Stunden die Woche arbeitest du?**

0-5

6-10

11-15

16-20

21-25

Sonstiges

Sonstiges

**B13. Fährst du eine ehrenamtliche Tätigkeit aus? (Verein, Hilfsorganisation, ...)**

Ja

Nein

**B14. Wie viele Stunden in der Woche führst du die Tätigkeit aus?**

0-3

4-7

8-11

12-15

16-19

20-23

24-27

28-31

32-35

36-39

40 und mehr

Sonstiges

Sonstiges

**B15. Hat deine ehrenamtliche Tätigkeit etwas mit deinem Studium zu tun?**

Ja

Nein

**B16. Hast du familiäre Verpflichtungen die sehr zeitintensiv sind?**

Auf Geschwister aufpassen

Pflegefall in der Familie

Kinderbetreuung

Nein

Sonstiges

Sonstiges

**B17. Bist du religiös?**

Ja

Nein

---

**B18. Benutzt du ein oder mehrere Social-Media-Kanäle?**

Facebook

Instagram

Twitter

TicToc

WhatsApp

Discord

Snapchat

Netflix

Prime Video

Disney plus

Sonstiges

Sonstiges

**B19. Wie viele Stunden verbringst du in der Woche mit Social-Media-Kanälen?**

0-3

4-7

8-11

12-15

16-19

20-23

24-27

28-31

32-35

36-39

40 und mehr

---

**B20. Bist du technisch interessiert?**

Sehr interessiert

Ein bisschen interessiert

Gar nicht interessiert

---

**Teil C: Fragen zu Prüfungen**

**C1. Hast du an digitalen Prüfungen teilgenommen?**

Ja

Nein

---

**C2. Wenn Ja, an welchen digitalen Prüfungsformen hast du teilgenommen?**

Take Home Exam als Moodle Abgabe

Take Home Exam als Moodle Test

Pool EDV Raum Prüfung

Schriftliche Prüfung über Zoom (mit Kamera)

Mündliche Zoom Prüfung

Sonstiges

Sonstiges

**C3. Besitzt du die nötige technische Ausstattung, um an online Prüfungen ohne Probleme teilnehmen zu können?**

Ich habe alle nötigen technischen Ausstattungen

Ich habe nicht alles, kann aber gut an Prüfungen teilnehmen

Mir fehlen die technischen Ausstattungen und ich kann nur bedingt an online Prüfungen teilnehmen

**C4. Bei welcher Prüfungsform hast du besser abgeschnitten?**

Online

Präsenz

Beide gleich

Noch keinen Vergleich

**C5. Bei welcher Prüfungsform fühlst du dich wohler?**

Online

Präsenz

Beide gleich

Noch keinen Vergleich

**C6. Welche Prüfungsform findest du fairer?**

Online

Präsenz

Beide gleich

Noch keinen Vergleich

**Teil D: Fragen zum Schummeln**

**D1. Wie stehst du zum Schummeln?**

Verwerflich wenn es andere machen

Neutrale Haltung

Wenn es meine Freunde machen ist es ok

Ich finde schummeln ok

Sonstiges

Sonstiges

**D2. Sind dir die Konsequenzen und Sanktionen des Schummeln in Klausuren bekannt?**

Ja

Nein

**D3. Welche Konsequenzen würden dich vom Schummeln abhalten?**

Mündliche Aufklärung vor der Prüfung über mögliche Konsequenzen

Mündliche persönliche Abmahnung während der Prüfung von der Prüfungsaufsicht

Notiz der Aufsicht (beim Erwischen), die der Prüfungskommission vorgelegt wird

Der Ausschluss der Prüfung (man darf nicht weiterschreiben)

Das nicht bestehen der Prüfung

Die Exmatrikulation

**D4. In welchem Versuch würdest du am ehesten Schummeln?**

1. Versuch

2. Versuch

3. Versuch

**D5. Beeinflusst der Dozent deine Bereitschaft zum Schummeln?**

**Ja = ich würde schummeln**

**Nein = ich würde nicht schummeln**

**große Vorlesung = viele Studenten, eher anonym**

**kleine Vorlesung = wenig Studenten**

	Ja	Unsicher	Nein
Sehr strenger Professor, große Vorlesung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sehr strenger Professor, kleine Vorlesung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Normaler Professor, große Vorlesung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Normaler Professor, kleine Vorlesung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sehr lockerer Professor, große Vorlesung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sehr lockerer Professor, kleine Vorlesung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**D6. Welche Schummel Methoden sind dir bekannt?**

Analoger Spickzettel

Manipulierte Prüfungsmaterialien

Einblendung von Inhalten auf den Hauptbildschirm ohne Wissen des Prüfers

Einblendung von Inhalten auf dem zweiten Bildschirm ohne Wissen des Prüfers

Einblendung von Inhalten auf einem anderen Gerät

Virtuelle Kamera: Studenten nehmen vor der Klausur ein Video aus der Perspektive ihrer Webcam von der vermeintlichen Prüfungssituation auf. Das Video wird in der realen Prüfung anstelle des Kamerabilds eingespielt.

Audio Signal im Ohr

Vortäuschung technischer Probleme

Vorbereitete Texte werden abgelesen

Übersetzungsprogramme

Kommunikation mit anderen Kommilitonen während der Prüfung

Lösungen von Dritten werden komplett kopiert

Prüfung wird von jemand anderen geschrieben (Andere Person sitzt im Raum, übernimmt die Steuerung des Computers, Prüfung wird über Cloud-Server geteilt)

Sonstiges

**D7. Welche Schummel Methode würdest du bei welcher Prüfung anwenden?Analoger Spickzettel**

Take Home Exam als Moodle Abgabe

Take Home Exam als Moodle Test

Pool EDV Raum Prüfung

Schriftliche Prüfung über Zoom (mit Kamera)

Mündliche Zoom Prüfung

**D8. Welche Schummel Methode würdest du bei welcher Prüfung anwenden?Manipulierte Prüfungsmaterialien**

Take Home Exam als Moodle Abgabe

Take Home Exam als Moodle Test

Pool EDV Raum Prüfung

Schriftliche Prüfung über Zoom (mit Kamera)

Mündliche Zoom Prüfung

**D9. Welche Schummel Methode würdest du bei welcher Prüfung anwenden?Einblendung von Inhalten auf den Hauptbildschirm ohne Wissen des Prüfers**

Take Home Exam als Moodle Abgabe

Take Home Exam als Moodle Test

Pool EDV Raum Prüfung

Schriftliche Prüfung über Zoom (mit Kamera)

Mündliche Zoom Prüfung

**D10. Welche Schummel Methode würdest du bei welcher Prüfung anwenden?Einblendung von Inhalten auf dem zweiten Bildschirm ohne Wissen des Prüfers**

Take Home Exam als Moodle Abgabe

Take Home Exam als Moodle Test

Pool EDV Raum Prüfung

	Schriftliche Prüfung über Zoom (mit Kamera)	<input type="checkbox"/>
	Mündliche Zoom Prüfung	<input type="checkbox"/>
<b>D11.</b>	<b>Welche Schummel Methode würdest du bei welcher Prüfung anwenden? Einblendung von Inhalten auf einem anderen Gerät</b>	
	Take Home Exam als Moodle Abgabe	<input type="checkbox"/>
	Take Home Exam als Moodle Test	<input type="checkbox"/>
	Pool EDV Raum Prüfung	<input type="checkbox"/>
	Schriftliche Prüfung über Zoom (mit Kamera)	<input type="checkbox"/>
	Mündliche Zoom Prüfung	<input type="checkbox"/>
<b>D12.</b>	<b>Welche Schummel Methode würdest du bei welcher Prüfung anwenden? Virtuelle Kamera: Studenten nehmen vor der Klausur ein Video aus der Perspektive ihrer Webcam von der vermeintlichen Prüfungssituation auf. Das Video wird in der realen Prüfung anstelle des Kamerabilds eingespielt.</b>	
	Take Home Exam als Moodle Abgabe	<input type="checkbox"/>
	Take Home Exam als Moodle Test	<input type="checkbox"/>
	Pool EDV Raum Prüfung	<input type="checkbox"/>
	Schriftliche Prüfung über Zoom (mit Kamera)	<input type="checkbox"/>
	Mündliche Zoom Prüfung	<input type="checkbox"/>
<b>D13.</b>	<b>Welche Schummel Methode würdest du bei welcher Prüfung anwenden? Audio Signal im Ohr</b>	
	Take Home Exam als Moodle Abgabe	<input type="checkbox"/>
	Take Home Exam als Moodle Test	<input type="checkbox"/>
	Pool EDV Raum Prüfung	<input type="checkbox"/>
	Schriftliche Prüfung über Zoom (mit Kamera)	<input type="checkbox"/>
	Mündliche Zoom Prüfung	<input type="checkbox"/>
<b>D14.</b>	<b>Welche Schummel Methode würdest du bei welcher Prüfung anwenden? Vortäuschung technischer Probleme</b>	
	Take Home Exam als Moodle Abgabe	<input type="checkbox"/>
	Take Home Exam als Moodle Test	<input type="checkbox"/>
	Pool EDV Raum Prüfung	<input type="checkbox"/>
	Schriftliche Prüfung über Zoom (mit Kamera)	<input type="checkbox"/>
	Mündliche Zoom Prüfung	<input type="checkbox"/>

<b>D15.</b>	<b>Welche Schummel Methode würdest du bei welcher Prüfung anwenden? Vorbereitete Texte werden abgelesen</b>	
	Take Home Exam als Moodle Abgabe	<input type="checkbox"/>
	Take Home Exam als Moodle Test	<input type="checkbox"/>
	Pool EDV Raum Prüfung	<input type="checkbox"/>
	Schriftliche Prüfung über Zoom (mit Kamera)	<input type="checkbox"/>
	Mündliche Zoom Prüfung	<input type="checkbox"/>
<b>D16.</b>	<b>Welche Schummel Methode würdest du bei welcher Prüfung anwenden? Übersetzungsprogramme</b>	
	Take Home Exam als Moodle Abgabe	<input type="checkbox"/>
	Take Home Exam als Moodle Test	<input type="checkbox"/>
	Pool EDV Raum Prüfung	<input type="checkbox"/>
	Schriftliche Prüfung über Zoom (mit Kamera)	<input type="checkbox"/>
	Mündliche Zoom Prüfung	<input type="checkbox"/>
<b>D17.</b>	<b>Welche Schummel Methode würdest du bei welcher Prüfung anwenden? Kommunikation mit anderen Kommilitonen während der Prüfung</b>	
	Take Home Exam als Moodle Abgabe	<input type="checkbox"/>
	Take Home Exam als Moodle Test	<input type="checkbox"/>
	Pool EDV Raum Prüfung	<input type="checkbox"/>
	Schriftliche Prüfung über Zoom (mit Kamera)	<input type="checkbox"/>
	Mündliche Zoom Prüfung	<input type="checkbox"/>
<b>D18.</b>	<b>Welche Schummel Methode würdest du bei welcher Prüfung anwenden? Lösungen von Dritten werden komplett kopiert</b>	
	Take Home Exam als Moodle Abgabe	<input type="checkbox"/>
	Take Home Exam als Moodle Test	<input type="checkbox"/>
	Pool EDV Raum Prüfung	<input type="checkbox"/>
	Schriftliche Prüfung über Zoom (mit Kamera)	<input type="checkbox"/>
	Mündliche Zoom Prüfung	<input type="checkbox"/>
<b>D19.</b>	<b>Welche Schummel Methode würdest du bei welcher Prüfung anwenden? Prüfung wird von jemand anderen geschrieben (Andere Person sitzt im Raum, übernimmt die Steuerung des Computers, Prüfung wird über Cloud-Server geteilt)</b>	
	Take Home Exam als Moodle Abgabe	<input type="checkbox"/>

	Take Home Exam als Moodle Test	<input type="checkbox"/>
	Pool EDV Raum Prüfung	<input type="checkbox"/>
	Schriftliche Prüfung über Zoom (mit Kamera)	<input type="checkbox"/>
	Mündliche Zoom Prüfung	<input type="checkbox"/>
<b>D20.</b>	<b>Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Analoger Spickzettel</b>	
	Mathe Aufgabe	<input type="checkbox"/>
	Multiple Choise	<input type="checkbox"/>
	Single Choise	<input type="checkbox"/>
	Offene Frage	<input type="checkbox"/>
	Transferfrage	<input type="checkbox"/>
	Definition	<input type="checkbox"/>
<b>D21.</b>	<b>Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Manipulierte Prüfungsmaterialien</b>	
	Mathe Aufgabe	<input type="checkbox"/>
	Multiple Choise	<input type="checkbox"/>
	Single Choise	<input type="checkbox"/>
	Offene Frage	<input type="checkbox"/>
	Transferfrage	<input type="checkbox"/>
	Definition	<input type="checkbox"/>
<b>D22.</b>	<b>Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Einblendung von Inhalten auf den Hauptbildschirm ohne Wissen des Prüfers</b>	
	Mathe Aufgabe	<input type="checkbox"/>
	Multiple Choise	<input type="checkbox"/>
	Single Choise	<input type="checkbox"/>
	Offene Frage	<input type="checkbox"/>
	Transferfrage	<input type="checkbox"/>
	Definition	<input type="checkbox"/>
<b>D23.</b>	<b>Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Einblendung von Inhalten auf dem zweiten Bildschirm ohne Wissen des Prüfers</b>	
	Mathe Aufgabe	<input type="checkbox"/>

	Multiple Choise	<input type="checkbox"/>
	Single Choise	<input type="checkbox"/>
	Offene Frage	<input type="checkbox"/>
	Transferfrage	<input type="checkbox"/>
	Definition	<input type="checkbox"/>
<b>D24.</b>	<b>Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Einblendung von Inhalten auf einem anderen Gerät</b>	
	Mathe Aufgabe	<input type="checkbox"/>
	Multiple Choise	<input type="checkbox"/>
	Single Choise	<input type="checkbox"/>
	Offene Frage	<input type="checkbox"/>
	Transferfrage	<input type="checkbox"/>
	Definition	<input type="checkbox"/>
<b>D25.</b>	<b>Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Virtuelle Kamera: Studenten nehmen vor der Klausur ein Video aus der Perspektive ihrer Webcam von der vermeintlichen Prüfungssituation auf. Das Video wird in der realen Prüfung anstelle des Kamerabilds eingespielt.</b>	
	Mathe Aufgabe	<input type="checkbox"/>
	Multiple Choise	<input type="checkbox"/>
	Single Choise	<input type="checkbox"/>
	Offene Frage	<input type="checkbox"/>
	Transferfrage	<input type="checkbox"/>
	Definition	<input type="checkbox"/>
<b>D26.</b>	<b>Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Audio Signal im Ohr</b>	
	Mathe Aufgabe	<input type="checkbox"/>
	Multiple Choise	<input type="checkbox"/>
	Single Choise	<input type="checkbox"/>
	Offene Frage	<input type="checkbox"/>
	Transferfrage	<input type="checkbox"/>
	Definition	<input type="checkbox"/>

**D27. Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Vortäuschung technischer Probleme**

Mathe Aufgabe

Multiple Choise

Single Choise

Offene Frage

Transferfrage

Definition

**D28. Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Vorbereitete Texte werden abgelesen**

Mathe Aufgabe

Multiple Choise

Single Choise

Offene Frage

Transferfrage

Definition

**D29. Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Übersetzungsprogramme**

Mathe Aufgabe

Multiple Choise

Single Choise

Offene Frage

Transferfrage

Definition

**D30. Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Kommunikation mit anderen Kommilitonen während der Prüfung**

Mathe Aufgabe

Multiple Choise

Single Choise

Offene Frage

Transferfrage

Definition

**D31. Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Lösungen von Dritten werden komplett kopiert**

Mathe Aufgabe

Multiple Choise

Single Choise

Offene Frage

Transferfrage

Definition

**D32. Bei welchem Aufgabentyp würdest deiner Meinung nach welche Schummel Methode passen?Prüfung wird von jemand anderen geschrieben (Andere Person sitzt im Raum, übernimmt die Steuerung des Computers, Prüfung wird über Cloud-Server geteilt)**

Mathe Aufgabe

Multiple Choise

Single Choise

Offene Frage

Transferfrage

Definition

**D33. Bei welcher digitalen Prüfungsform hast du schon einmal geschummelt?**

Take Home Exam als Moodle Abgabe

Take Home Exam als Moodle Test

Pool EDV Raum Prüfung

Schriftliche Prüfung über Zoom (mit Kamera)

Mündliche Zoom Prüfung

**D34. Wie schummelst du?**



## 2. Auszug aus der bearbeiteten Exceldatei mit den Umfrageergebnissen

Studenten ID	Studiengang	Fachsemester	Notendurchschnitt	Alter	Geschlecht	Wohnort
1	1 Wirtschaftsingenieurwesen	3	3,25	23	Weiblich	Bei deinen Eltern
2	2 Wirtschaftsingenieurwesen	1	2,14	19	männlich	Bei deinen Eltern
3	3 Wirtschaftsingenieurwesen	1	2,14	18	männlich	Bei deinen Eltern
4	4 Wirtschaftsingenieurwesen	1	2,14	19	männlich	Bei deinen Eltern
5	5 Wirtschaftsingenieurwesen	1	3,25	23	männlich	Bei deinen Eltern
6	6 Wirtschaftsingenieurwesen	1	2,14	20	männlich	Bei deinen Eltern
7	7 Wirtschaftsingenieurwesen	1	2,14	19	männlich	Eigene Wohnung
8	8 Wirtschaftsingenieurwesen	1	2,14	nicht angegeben		Bei deinen Eltern
9	9 Wirtschaftsingenieurwesen	1	2,14	19	männlich	WG-Zimmer
10	10 Wirtschaftsingenieurwesen	1	2,14	19	Weiblich	Bei deinen Eltern
11	11 Wirtschaftsingenieurwesen	1	2,14	22	männlich	Bei deinen Eltern
12	12 Wirtschaftsingenieurwesen	1	2,14	24	männlich	WG-Zimmer
13	13 Wirtschaftsingenieurwesen	2	2,25		männlich	Eigene Wohnung
14	14 Wirtschaftsingenieurwesen	1	2,14	20	männlich	Bei deinen Eltern
15	15 Wirtschaftsingenieurwesen	1	2,14	22	männlich	Bei deinen Eltern
16	16 Wirtschaftsingenieurwesen	1	2,14	26	männlich	Eigene Wohnung
17	17 Wirtschaftsingenieurwesen	1	2,14	20	männlich	Sonstiges
18	18 Wirtschaftsingenieurwesen	1	2,14		männlich	Bei deinen Eltern
19	19 Wirtschaftsingenieurwesen	1	2,14	19	männlich	Studentenwohnheim
20	20 Wirtschaftsingenieurwesen	1	2,14		Weiblich	Bei deinen Eltern
21	21 Wirtschaftsingenieurwesen	1	2,14	nicht angegeben		
22	22 Informationsmanagement im Gesundheitswesen	1	1,75	19	Weiblich	Studentenwohnheim
23	23 Betriebswirtschaft	4	1,75	nicht angegeben		
24	24 Betriebswirtschaft im Gesundheitswesen	9	2,25	39	männlich	Eigene Wohnung
25	25 Sonstiges	3	1,25	23	Weiblich	Eigene Wohnung
26	26 Informationsmanagement und Unternehmenskommunikation	4	1,75	24	Weiblich	Studentenwohnheim
27	27 Informationsmanagement und Unternehmenskommunikation	7	1,25	27	Weiblich	Bei deinen Eltern
28	28 Sonstiges	2	1,25	25	männlich	Bei deinen Eltern
29	29 Wirtschaftspsychologie	6	1,75	23	Weiblich	WG-Zimmer
30	30 Sonstiges	1	2,14	24	Weiblich	WG-Zimmer
31	31 Informationsmanagement und Unternehmenskommunikation	3	2,75	24	männlich	WG-Zimmer
32	32 Wirtschaftsingenieurwesen	8	2,25	27	männlich	Eigene Wohnung
33	33	1	1,75	30	Weiblich	Eigene Wohnung
34	34 Data Science Management	3	1,75	20	Weiblich	Studentenwohnheim
35	35 Betriebswirtschaft	7	1,75	22	Weiblich	WG-Zimmer
36	36 Wirtschaftspsychologie	1	2,14	18	Weiblich	Studentenwohnheim
37	36 Wirtschaftspsychologie	1	2,14	18	Weiblich	Studentenwohnheim

### 3. Jupiter Notebook der deskriptiven Auswertung

#### Bachelorarbeit von Stefanie Wittgens

Untersuchung von Täuschungsverhalten in digitalen Prüfungen Identifikation von Einflussfaktoren mithilfe von maschinellem Lernen

#### Deskriptive Auswertung

Auswertung der demographischen Daten

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import plot_likert
```

```
In [2]: df = pd.read_excel(r"C:\Users\stefa\Desktop\BA_Umfrage.xlsx", sheet_name='Demographische Daten')
df.head()
```

Out[2]:

	Studenten ID	Studiengang	Fachsemester	Notendurchschnitt	Alter	Geschlecht	Wohnort
0	Student 1	Wirtschaftsingenieurwesen	3.0	3.250000	23.0	Weiblich	Bei deinen Eltern
1	Student 2	Wirtschaftsingenieurwesen	1.0	2.141753	19.0	männlich	Bei deinen Eltern
2	Student 3	Wirtschaftsingenieurwesen	1.0	2.141753	18.0	männlich	Bei deinen Eltern
3	Student 4	Wirtschaftsingenieurwesen	1.0	2.141753	19.0	männlich	Bei deinen Eltern
4	Student 5	Wirtschaftsingenieurwesen	1.0	3.250000	23.0	männlich	Bei deinen Eltern

```
In [3]: df.describe().T
```

Out[3]:

	count	mean	std	min	25%	50%	75%	max
Fachsemester	255.0	4.015686	2.517605	1.00	2.00	4.000000	6.00	11.00
Notendurchschnitt	257.0	2.141753	0.485037	1.25	1.75	2.141753	2.25	3.75
Alter	238.0	23.323529	4.822630	18.00	21.00	22.000000	24.00	56.00

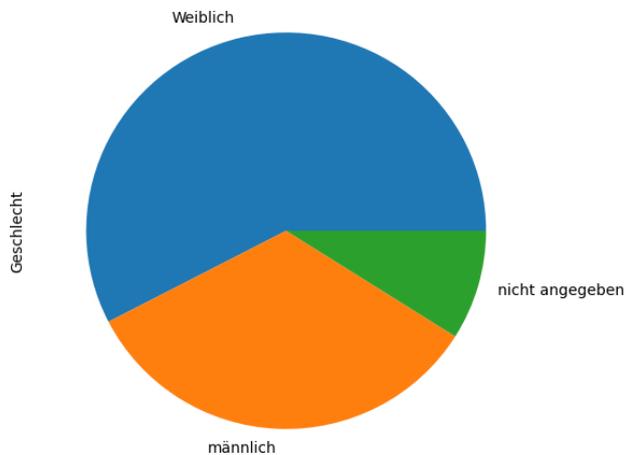
```
In [4]: df.Geschlecht.value_counts()
```

```
Out[4]: Weiblich      148
männlich      86
nicht angegeben  23
Name: Geschlecht, dtype: int64
```

```
In [5]: df.Geschlecht.value_counts(normalize=True)
```

```
Out[5]: Weiblich      0.575875
männlich      0.334638
nicht angegeben  0.089494
Name: Geschlecht, dtype: float64
```

```
In [6]: df.Geschlecht.value_counts().plot.pie(figsize=(6,6));
```



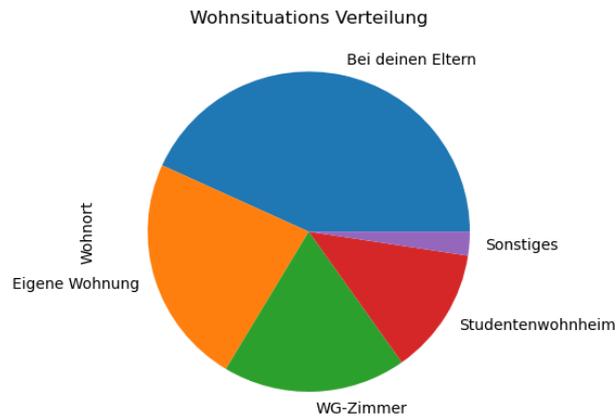
```
In [7]: df.Wohnort.value_counts()
```

```
Out[7]: Bei deinen Eltern    108  
Eigene Wohnung             58  
WG-Zimmer                  46  
Studentenwohnheim         32  
Sonstiges                   6  
Name: Wohnort, dtype: int64
```

```
In [8]: df.Wohnort.value_counts(normalize=True)
```

```
Out[8]: Bei deinen Eltern    0.432  
Eigene Wohnung             0.232  
WG-Zimmer                  0.184  
Studentenwohnheim         0.128  
Sonstiges                   0.024  
Name: Wohnort, dtype: float64
```

```
In [9]: df.Wohnort.value_counts().plot(kind='pie', title='Wohnsituations Verteilung');
```



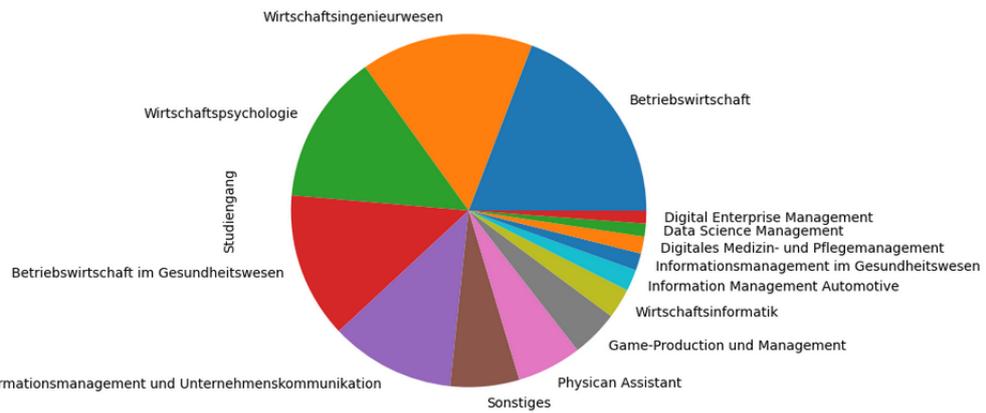
```
In [10]: df.Studiengang.value_counts()
```

```
Out[10]: Betriebswirtschaft                49  
Wirtschaftsingenieurwesen             40  
Wirtschaftspsychologie                 35  
Betriebswirtschaft im Gesundheitswesen 34  
Informationsmanagement und Unternehmenskommunikation 29  
Sonstiges                               16  
Physican Assistant                      15  
Game-Production und Management          11  
Wirtschaftsinformatik                   7  
Information Management Automotive        5  
Informationsmanagement im Gesundheitswesen 4  
Digitales Medizin- und Pflegemanagement 4  
Data Science Management                  3  
Digital Enterprise Management            3  
Name: Studiengang, dtype: int64
```

```
In [11]: df.Studiengang.value_counts(normalize=True)
```

```
Out[11]: Betriebswirtschaft                0.192157  
Wirtschaftsingenieurwesen             0.156863  
Wirtschaftspsychologie                 0.137255  
Betriebswirtschaft im Gesundheitswesen 0.133333  
Informationsmanagement und Unternehmenskommunikation 0.113725  
Sonstiges                               0.062745  
Physican Assistant                      0.058824  
Game-Production und Management          0.043137  
Wirtschaftsinformatik                   0.027451  
Information Management Automotive        0.019608  
Informationsmanagement im Gesundheitswesen 0.015686  
Digitales Medizin- und Pflegemanagement 0.015686  
Data Science Management                  0.011765  
Digital Enterprise Management            0.011765  
Name: Studiengang, dtype: float64
```

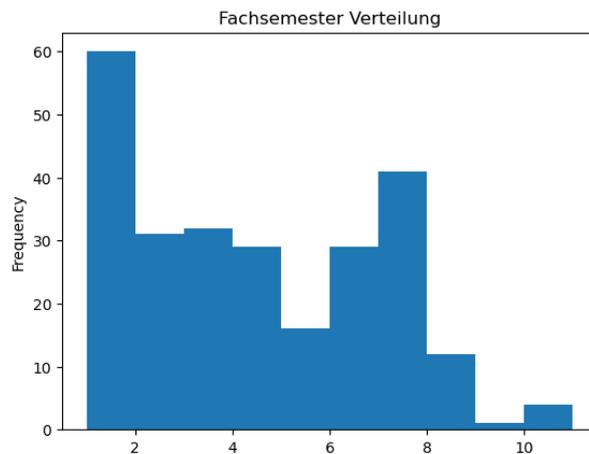
```
In [12]: df.Studiengang.value_counts().plot.pie(figsize=(6,6));
```



```
In [13]: df.Fachsemester.value_counts(normalize=True)
```

```
Out[13]: 1.0    0.235294
         7.0    0.160784
         3.0    0.125490
         2.0    0.121569
         4.0    0.113725
         6.0    0.113725
         5.0    0.062745
         8.0    0.047059
        11.0    0.015686
         9.0    0.003922
         Name: Fachsemester, dtype: float64
```

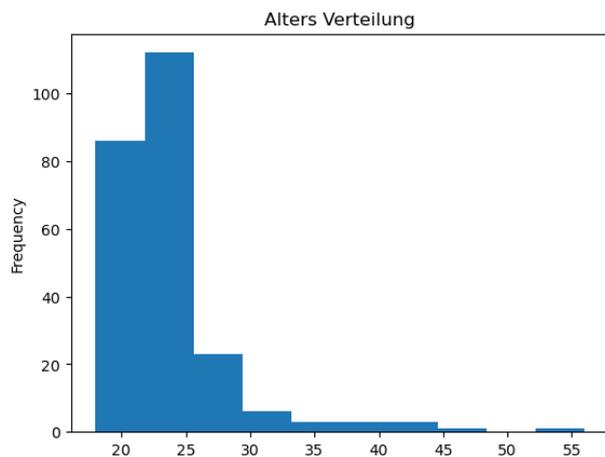
```
In [14]: df['Fachsemester'].plot(kind='hist', title='Fachsemester Verteilung');
```



```
In [15]: df.Alter.value_counts(normalize=True)
```

```
Out[15]: 22.0    0.172269
         23.0    0.147059
         21.0    0.130252
         20.0    0.126050
         24.0    0.092437
         19.0    0.067227
         25.0    0.058824
         27.0    0.042017
         26.0    0.037815
         18.0    0.037815
         31.0    0.012605
         30.0    0.012605
         41.0    0.008403
         29.0    0.008403
         28.0    0.008403
         39.0    0.004202
         35.0    0.004202
         34.0    0.004202
         56.0    0.004202
         36.0    0.004202
         43.0    0.004202
         40.0    0.004202
         37.0    0.004202
         46.0    0.004202
         Name: Alter, dtype: float64
```

```
In [16]: df['Alter'].plot(kind='hist', title='Alters Verteilung');
```



Auswertung der Zufriedenheit abgefragt über die Likert-Skala

```
In [17]: df1 = pd.read_excel(r"C:\Users\stefa\Desktop\BA_Umfrage.xlsx", sheet_name='Zufriedenheit')
df1.head()
```

Out[17]:

	Studenten ID	Studium	Hochschule	Motivation	Leistungsdruck
0	Student 1	Weder noch	Stimme zu	Unmotiviert	Stimme zu
1	Student 2	Stimme zu	Stimme nicht zu	Motiviert	Weder noch
2	Student 3	Stimme zu	Stimme zu	Motiviert	Stimme zu
3	Student 4	Stimme voll und ganz zu	Stimme zu	Sehr motiviert	Stimme zu
4	Student 5	Stimme zu	Stimme voll und ganz zu	Motiviert	Weder noch

```
In [18]: df1.describe()
```

Out[18]:

	Studenten ID	Studium	Hochschule	Motivation	Leistungsdruck
count	257	256	255	256	254
unique	257	5	5	5	5
top	Student 1	Stimme zu	Stimme zu	Motiviert	Stimme zu
freq	1	176	114	121	95

```
In [19]: df1.Studium.value_counts(normalize=True)
```

Out[19]:

Stimme zu	0.687500
Weder noch	0.140625
Stimme voll und ganz zu	0.089844
Stimme nicht zu	0.074219
Stimme überhaupt nicht zu	0.007812

Name: Studium, dtype: float64

```
In [20]: df1.Hochschule.value_counts(normalize=True)
```

Out[20]:

Stimme zu	0.447059
Stimme nicht zu	0.278431
Weder noch	0.172549
Stimme überhaupt nicht zu	0.054902
Stimme voll und ganz zu	0.047059

Name: Hochschule, dtype: float64

```
In [21]: df1.Motivation.value_counts(normalize=True)
```

Out[21]:

Motiviert	0.472656
Weder noch	0.339844
Unmotiviert	0.105469
Sehr motiviert	0.050781
Sehr unmotiviert	0.031250

Name: Motivation, dtype: float64

```
In [22]: df1.Leistungsdruck.value_counts(normalize=True)
```

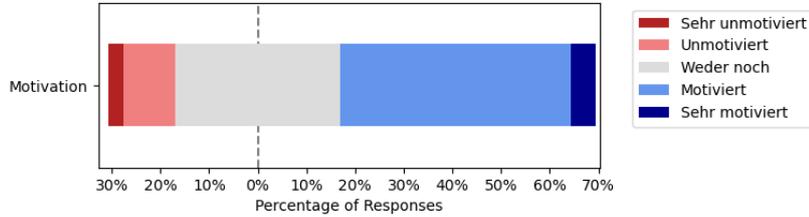
Out[22]:

Stimme zu	0.374016
Stimme voll und ganz zu	0.228346
Weder noch	0.196850
Stimme nicht zu	0.122047
Stimme überhaupt nicht zu	0.078740

Name: Leistungsdruck, dtype: float64

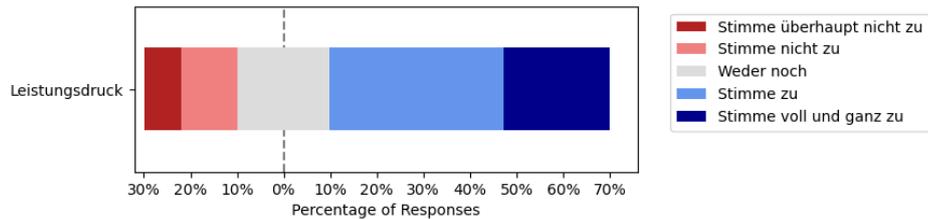
```
In [23]: scala2 = \
    ['Sehr unmotiviert',
     'Unmotiviert',
     'Weder noch',
     'Motiviert',
     'Sehr motiviert']

# create a likert plot
plot_likert.plot_likert(df1["Motivation"], scala2, plot_percentage=True, figsize = (6,2));
```



```
In [24]: scala1 = \
    ['Stimme überhaupt nicht zu',
     'Stimme nicht zu',
     'Weder noch',
     'Stimme zu',
     'Stimme voll und ganz zu']

plot_likert.plot_likert(df1["Leistungsdruck"], scala1, plot_percentage=True, figsize = (6,2));
```

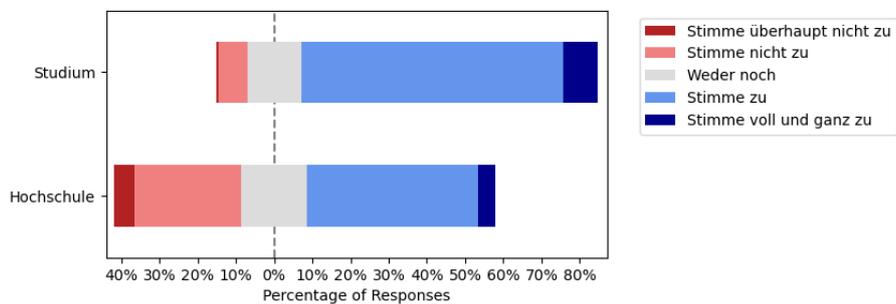


```
In [25]: df1_neu = df1.drop(columns=["Studenten ID", "Motivation", "Leistungsdruck"])

scala1 = \
    ['Stimme überhaupt nicht zu',
     'Stimme nicht zu',
     'Weder noch',
     'Stimme zu',
     'Stimme voll und ganz zu']

plot_likert.plot_likert(df1_neu, scala1, plot_percentage=True, figsize = (6,3));
```

C:\Users\stefa\anaconda3\lib\site-packages\plot\_likert\plot\_likert.py:299: UserWarning: In your data, not all questions have the same number of responses. i.e., different numbers of people answered each question. Therefore, the percentages aren't directly comparable: X% for one question represents a different number of responses than X% for another question, yet they will appear the same in the percentage graph. This may be misleading to your reader.



Auswertung der Hobbies und Sportarten

```
In [26]: df2 = pd.read_excel(r"C:\Users\stefa\Desktop\BA_Umfrage.xlsx", sheet_name='Hobbies')
df2.head()
```

Out[26]:

Studenten ID	Fußball	Basketball	Tennis	Tanzen	Yoga	Schwimmen	Fitnessstudio	Pilates	Klettern	...	Joggen/Laufen	Schach	Poker	Freunde Treffen	Lesen	
0 Student 1	Nein	Nein	Nein	Nein	Nein	Nein	Nein	Ja	Nein	Nein	...	Nein	Nein	Nein	Ja	Ja
1 Student 2	Nein	Nein	Nein	Nein	Nein	Nein	Nein	Nein	Nein	Nein	...	Nein	Nein	Nein	Ja	Nein
2 Student 3	Ja	Ja	Nein	Nein	Nein	Nein	Nein	Ja	Nein	Nein	...	Nein	Nein	Nein	Nein	Nein
3 Student 4	Nein	Nein	Ja	Nein	Nein	Nein	Nein	Nein	Nein	Ja	...	Ja	Ja	Ja	Ja	Nein
4 Student 5	Nein	Nein	Nein	Nein	Nein	Nein	Nein	Ja	Nein	Nein	...	Nein	Nein	Nein	Ja	Nein

5 rows x 23 columns

```
In [27]: # Umwandlung der Ja's und Nein's in Zahlenwerte zur besseren Auswertung
df2 = df2.replace(["Ja"], [1])
df2 = df2.replace(["Nein"], [0])
df2.head()
```

Out[27]:

Studenten ID	Fußball	Basketball	Tennis	Tanzen	Yoga	Schwimmen	Fitnessstudio	Pilates	Klettern	...	Joggen/Laufen	Schach	Poker	Freunde Treffen	Lesen	
0 Student 1	0	0	0	0	0	0	0	1	0	0	...	0	0	0	1	1
1 Student 2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0
2 Student 3	1	1	0	0	0	0	0	1	0	0	...	0	0	0	0	0
3 Student 4	0	0	1	0	0	0	0	0	0	1	...	1	1	1	1	0
4 Student 5	0	0	0	0	0	0	0	1	0	0	...	0	0	0	1	0

5 rows x 23 columns

```
In [28]: df2.describe()
```

Out[28]:

	Fußball	Basketball	Tennis	Tanzen	Yoga	Schwimmen	Fitnessstudio	Pilates	Klettern	Kampfsport	...	Handball	Jo
count	257.000000	257.000000	257.000000	257.000000	257.000000	257.000000	257.000000	257.000000	257.000000	257.000000	...	257.000000	
mean	0.124514	0.038911	0.054475	0.108949	0.070039	0.085603	0.408560	0.066148	0.062257	0.015564	...	0.015564	
std	0.330811	0.193759	0.227395	0.312184	0.255711	0.280323	0.492527	0.249025	0.242093	0.124023	...	0.124023	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	
75%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	...	0.000000	
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000	

8 rows x 21 columns

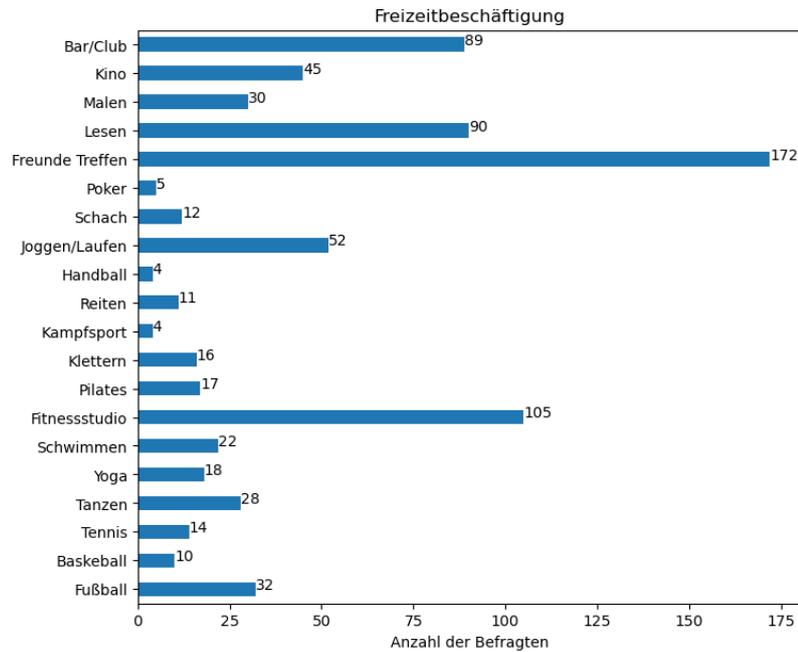
```
In [29]: filt_df2=df2.loc[:,df2.columns!="Studenten ID"]
filtered_df2=filt_df2.loc[:,filt_df2.columns!="Zeit_Hobbies/Sport"]
#filtered_df2.head()
filtered_df2.sum(numeric_only=True)
```

Out[29]:

```
Fußball          32
Basketball       10
Tennis           14
Tanzen           28
Yoga             18
Schwimmen        22
Fitnessstudio    105
Pilates          17
Klettern         16
Kampfsport        4
Reiten           11
Handball          4
Joggen/Laufen    52
Schach           12
Poker            5
Freunde Treffen  172
Lesen            90
Malen            30
Kino             45
Bar/Club         89
dtype: int64
```

```
In [30]: plt.figure(figsize=(8,7))
filtered_df2.sum(numeric_only=True).plot.barh();
# Loop über die Bars und hinzufügen der Werte als Beschriftungen
for i, v in enumerate(filtered_df2.sum(numeric_only=True).values):
    plt.text(v, i, str(v))

plt.title("Freizeitbeschäftigung");
plt.xlabel("Anzahl der Befragten");
```



#### Auswertung der Schulischen Aktivitäten

```
In [31]: df3 = pd.read_excel(r"C:\Users\stefa\Desktop\BA_Umfrage.xlsx", sheet_name='Schulische Aktivitäten')
df3.head()
```

```
Out[31]:
```

	Studenten ID	Vor-Nachbearbeitung	Regelmäßig_Vorlesung
0	Student 1	2.5	Ich lasse regelmäßig Vorlesungen ausfallen
1	Student 2	2.5	Ich besuche fast alle Vorlesungen
2	Student 3	17.5	Ich besuche alle Vorlesungen
3	Student 4	7.5	Ich besuche alle Vorlesungen
4	Student 5	7.5	Ich besuche alle Vorlesungen

```
In [32]: df3.describe().T
```

```
Out[32]:
```

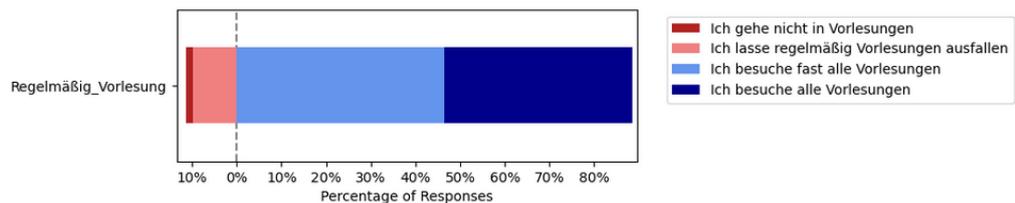
	count	mean	std	min	25%	50%	75%	max
Vor-Nachbearbeitung	246.0	12.317073	8.902366	2.5	7.5	10.0	17.5	37.5

```
In [33]: df3.Regelmäßig_Vorlesung.value_counts(normalize=True)
```

```
Out[33]: Ich besuche fast alle Vorlesungen      0.464567
Ich besuche alle Vorlesungen                  0.421260
Ich lasse regelmäßig Vorlesungen ausfallen    0.098425
Ich gehe nicht in Vorlesungen                 0.015748
Name: Regelmäßig_Vorlesung, dtype: float64
```

```
In [34]: scala3 = \
    ['Ich gehe nicht in Vorlesungen',
     'Ich lasse regelmäßig Vorlesungen ausfallen',
     '..',
     'Ich besuche fast alle Vorlesungen',
     'Ich besuche alle Vorlesungen',]

plot_likert.plot_likert(df3["Regelmäßig_Vorlesung"], scala3, plot_percentage=True, figsize = (6,2));
```



Auswertung des Verhaltens in Prüfungen

```
In [35]: df4 = pd.read_excel(r"C:\Users\stefa\Desktop\BA_Umfrage.xlsx", sheet_name='Prüfungsverhalten')
df4.head()
```

Out[35]:

	Studenten ID	Teilgenommen	technische_Ausstattung	Prüfungsform_besser	Prüfungsform_Wohlfühlen	Prüfungsform_fairer
0	Student 1	NaN	Ich habe alle nötigen technischen Ausstattungen	Präsenz	Beide gleich	Beide gleich
1	Student 2	Ja	Ich habe alle nötigen technischen Ausstattungen	Noch keinen Vergleich	Online	Online
2	Student 3	Nein	Ich habe alle nötigen technischen Ausstattungen	Noch keinen Vergleich	Beide gleich	Beide gleich
3	Student 4	NaN	Ich habe alle nötigen technischen Ausstattungen	Noch keinen Vergleich	Online	Beide gleich
4	Student 5	Nein	NaN	Noch keinen Vergleich	Noch keinen Vergleich	Noch keinen Vergleich

```
In [36]: df4.Teilgenommen.value_counts(normalize=True)
```

Out[36]:

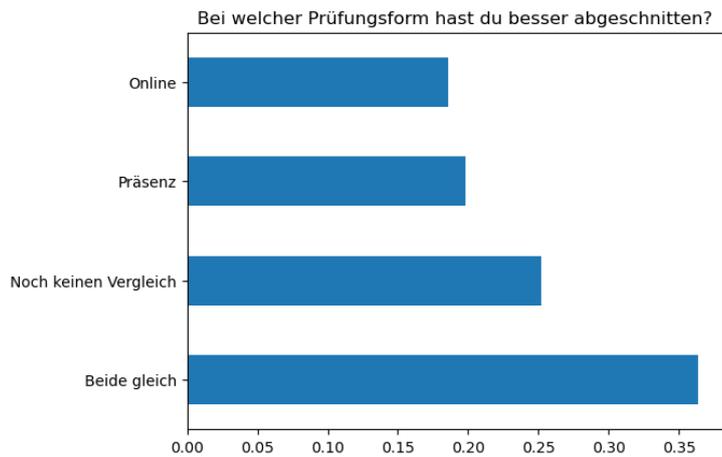
```
Ja      0.788018
Nein    0.211982
Name: Teilgenommen, dtype: float64
```

```
In [37]: df4.Prüfungsform_besser.value_counts(normalize=True)
```

Out[37]:

```
Beide gleich      0.363636
Noch keinen Vergleich  0.252066
Präsenz           0.198347
Online            0.185950
Name: Prüfungsform_besser, dtype: float64
```

```
In [38]: df4.Prüfungsform_besser.value_counts(normalize=True).plot.barh();
plt.title("Bei welcher Prüfungsform hast du besser abgeschnitten?");
```

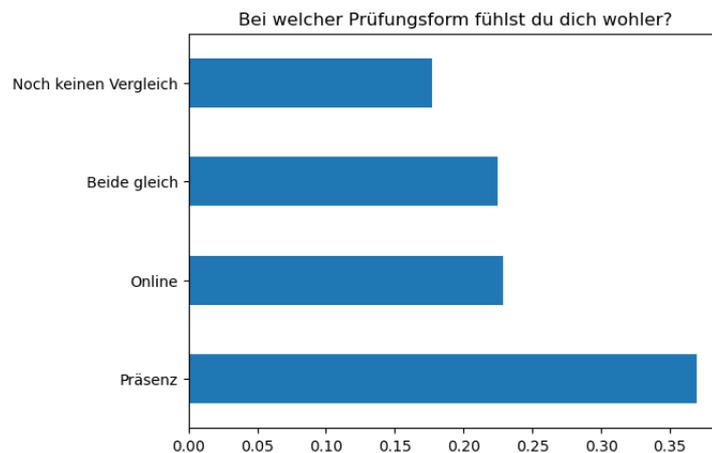


```
In [39]: df4.Prüfungsform_Wohlfühlen.value_counts(normalize=True)
```

Out[39]:

```
Präsenz      0.369478
Online       0.228916
Beide gleich  0.224900
Noch keinen Vergleich  0.176707
Name: Prüfungsform_Wohlfühlen, dtype: float64
```

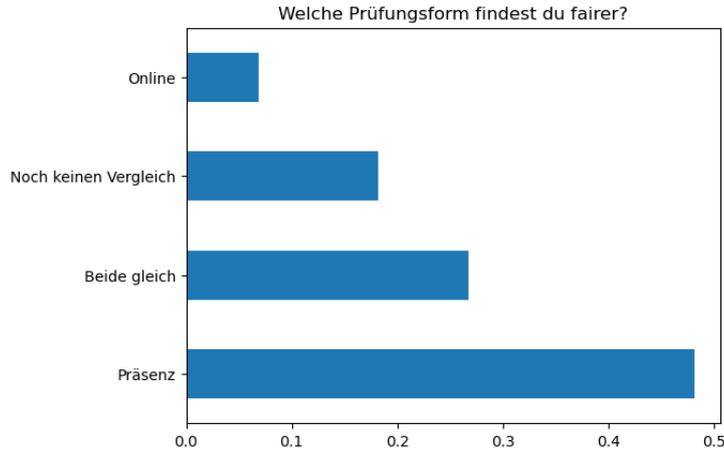
```
In [40]: df4.Prüfungsform_Wohlfühlen.value_counts(normalize=True).plot.barh();
plt.title("Bei welcher Prüfungsform fühlst du dich wohler?");
```



```
In [41]: df4.Prüfungsform_fairer.value_counts(normalize=True)
```

```
Out[41]: Präsenz          0.481781  
Beide gleich          0.267206  
Noch keinen Vergleich  0.182186  
Online                0.068826  
Name: Prüfungsform_fairer, dtype: float64
```

```
In [42]: df4.Prüfungsform_fairer.value_counts(normalize=True).plot.barh();  
plt.title("Welche Prüfungsform findest du fairer?");
```



```
In [43]: df5 = pd.read_excel(r"C:\Users\stefa\Desktop\BA_Umfrage.xlsx", sheet_name='DP Teilnahmen')  
df5.head()
```

```
Out[43]:
```

	Studenten ID	Take_Home_Moodle_Abgabe	Take_Home_Moodle_Test	Pool_EDV_Raum_Prüfung	Schriftliche_Prüfung_Zoom_(mit Kamera)	Mündliche_Zoom_Prüfung	So
0	Student 1	Ja	Nein	Nein	Nein	Nein	
1	Student 2	Ja	Ja	Ja	Nein	Nein	
2	Student 3	Nein	Nein	Nein	Nein	Nein	
3	Student 4	Ja	Nein	Ja	Nein	Nein	
4	Student 5	Nein	Nein	Nein	Nein	Nein	

```
In [44]: # Umwandlung der Ja's und Nein's in Zahlenwerte zur besseren Auswertung  
df5 = df5.replace(["Ja"], [1])  
df5 = df5.replace(["Nein"], [0])  
df5.head()
```

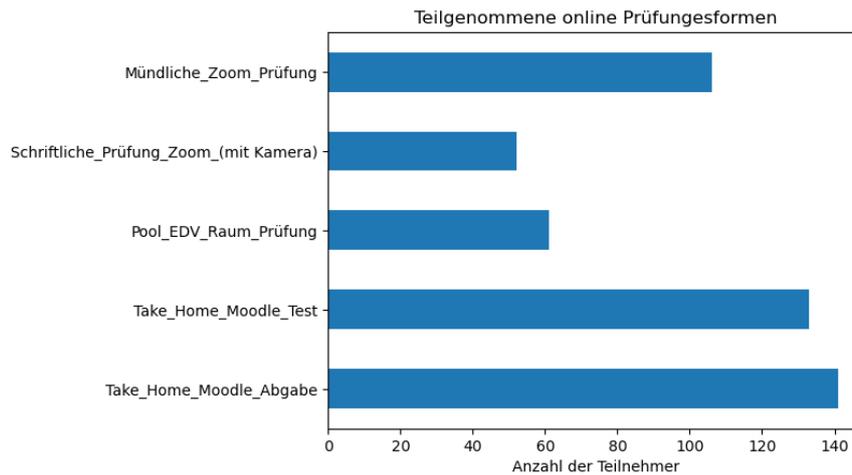
```
Out[44]:
```

	Studenten ID	Take_Home_Moodle_Abgabe	Take_Home_Moodle_Test	Pool_EDV_Raum_Prüfung	Schriftliche_Prüfung_Zoom_(mit Kamera)	Mündliche_Zoom_Prüfung	So
0	Student 1	1	0	0	0	0	
1	Student 2	1	1	1	0	0	
2	Student 3	0	0	0	0	0	
3	Student 4	1	0	1	0	0	
4	Student 5	0	0	0	0	0	

```
In [45]: filtered_df5=df5.loc[:,df5.columns!="Studenten ID"];  
filtered_df5.sum(numeric_only=True)
```

```
Out[45]: Take_Home_Moodle_Abgabe          141  
Take_Home_Moodle_Test          133  
Pool_EDV_Raum_Prüfung           61  
Schriftliche_Prüfung_Zoom_(mit Kamera)  52  
Mündliche_Zoom_Prüfung          106  
dtype: int64
```

```
In [46]: filtered_df5.sum(numeric_only=True).plot.barh();
plt.title("Teilgenommene online Prüfungsformen");
plt.xlabel("Anzahl der Teilnehmer");
```



```
In [47]: df6 = pd.read_excel(r"C:\Users\stefa\Desktop\BA_Umfrage.xlsx", sheet_name='DP SchummeIn')
df6.head()
```

Out[47]:

	Studenten ID	Take_Home_Moodle_Abgabe	Take_Home_Moodle_Test	Pool_EDV_Raum_Prüfung	Schriftliche_Prüfung_Zoom_(mit Kamera)	Mündliche_Zoom_Prüfung
0	Student 1	Ja	Nein	Nein	Nein	Nein
1	Student 2	Nein	Nein	Nein	Nein	Nein
2	Student 3	Nein	Nein	Nein	Nein	Nein
3	Student 4	Nein	Nein	Nein	Nein	Nein
4	Student 5	Nein	Nein	Nein	Nein	Nein

```
In [48]: df6 = df6.replace(["Ja"], [1])
df6 = df6.replace(["Nein"], [0])
df6.head()
```

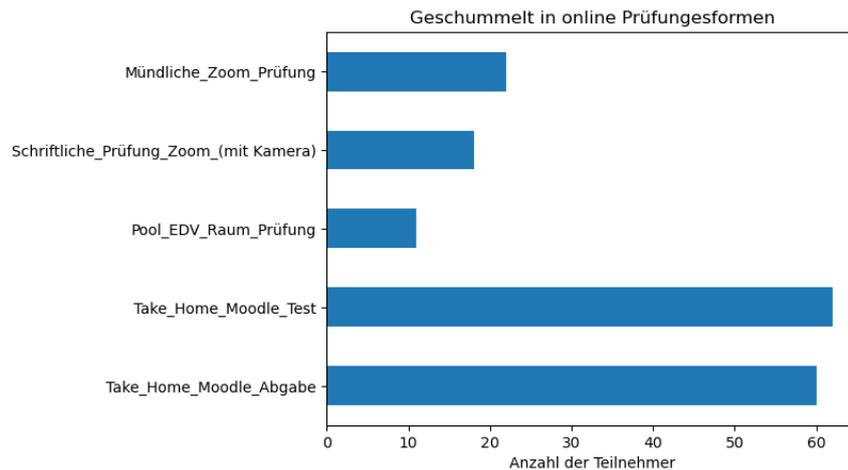
Out[48]:

	Studenten ID	Take_Home_Moodle_Abgabe	Take_Home_Moodle_Test	Pool_EDV_Raum_Prüfung	Schriftliche_Prüfung_Zoom_(mit Kamera)	Mündliche_Zoom_Prüfung
0	Student 1	1	0	0	0	0
1	Student 2	0	0	0	0	0
2	Student 3	0	0	0	0	0
3	Student 4	0	0	0	0	0
4	Student 5	0	0	0	0	0

```
In [49]: filtered_df6=df6.loc[:,df6.columns!="Studenten ID"];
filtered_df6.sum(numeric_only=True)
```

```
Out[49]: Take_Home_Moodle_Abgabe      60
Take_Home_Moodle_Test      62
Pool_EDV_Raum_Prüfung      11
Schriftliche_Prüfung_Zoom_(mit Kamera)  18
Mündliche_Zoom_Prüfung      22
dtype: int64
```

```
In [50]: filtered_df6.sum(numeric_only=True).plot.barh();
plt.title("Geschummelt in online Prüfungsformen");
plt.xlabel("Anzahl der Teilnehmer");
```



```
In [51]: df7 = pd.read_excel(r"C:\Users\stefa\Desktop\BA_Umfrage.xlsx", sheet_name='Schummelverhalten')
df7.head()
```

Out[51]:

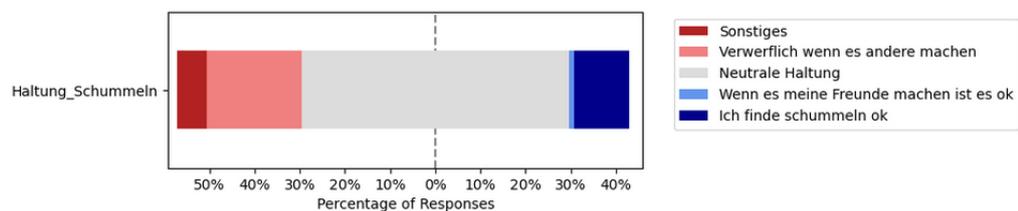
	Studenten ID	Haltung_Schummeln	Konsequenzen_bekannt	Konsequenzen_abhalten	Wie schummelst du?
0	Student 1	Neutrale Haltung	Ja	Die Exmatrikulation	Zettel
1	Student 2	Neutrale Haltung	Ja	Das nicht bestehen der Prüfung	In der Schule konnte man während Moodle Klasse...
2	Student 3	Verwerflich wenn es andere machen	Ja	NaN	NIE!
3	Student 4	Wenn es meine Freunde machen ist es ok	Ja	Die Exmatrikulation	nicht
4	Student 5	Neutrale Haltung	Ja	Das nicht bestehen der Prüfung	in der Regal gar nicht

```
In [52]: df7.Haltung_Schummeln.value_counts(normalize=True)
```

```
Out[52]: Neutrale Haltung          0.591093
Verwerflich wenn es andere machen  0.210526
Ich finde schummeln ok             0.121457
Sonstiges                          0.064777
Wenn es meine Freunde machen ist es ok 0.012146
Name: Haltung_Schummeln, dtype: float64
```

```
In [53]: scala4 = \
['Sonstiges',
'Verwerflich wenn es andere machen',
'Neutrale Haltung',
'Wenn es meine Freunde machen ist es ok',
'Ich finde schummeln ok']

plot_likert.plot_likert(df7["Haltung_Schummeln"], scala4, plot_percentage=True, figsize = (6,2));
```



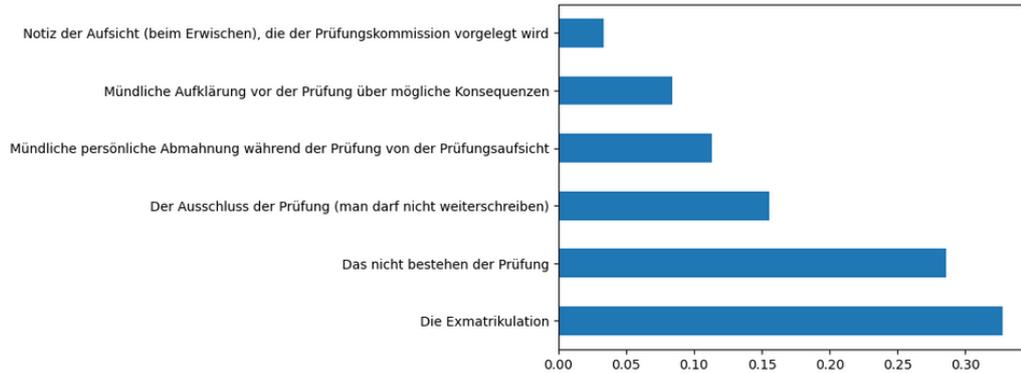
```
In [54]: df7.Konsequenzen_bekannt.value_counts(normalize=True)
```

```
Out[54]: Ja          0.896266
Nein       0.103734
Name: Konsequenzen_bekannt, dtype: float64
```

```
In [55]: df7.Konsequenzen_abhalten.value_counts(normalize=True)
```

```
Out[55]: Die Exmatrikulation          0.327731
Das nicht bestehen der Prüfung       0.285714
Der Ausschluss der Prüfung (man darf nicht weiterschreiben) 0.155462
Mündliche persönliche Abmahnung während der Prüfung von der Prüfungsaufsicht 0.113445
Mündliche Aufklärung vor der Prüfung über mögliche Konsequenzen 0.084034
Notiz der Aufsicht (beim Erwischen), die der Prüfungskommission vorgelegt wird 0.033613
Name: Konsequenzen_abhalten, dtype: float64
```

```
In [56]: df7.Konsequenzen_abhalten.value_counts(normalize=True).plot.barh();
```



```
In [57]: df8 = pd.read_excel(r"C:\Users\stefa\Desktop\BA_Umfrage.xlsx", sheet_name='Bekannte Schummelmethoden')
df8.head()
```

Out[57]:

	Studenten ID	Analoger Spickzettel	Manipulierte Prüfungsmaterialien	Einblendung von Inhalten auf den Hauptbildschirm ohne Wissen des Prüfers	Einblendung von Inhalten auf dem zweiten Bildschirm ohne Wissen des Prüfers	Einblendung von Inhalten auf einem anderen Gerät	Virtuelle Kamera	Audio Signal im Ohr	Vortäuschung technischer Probleme	Vorbereitete Texte werden abgelesen	Übersetzungsprogra
0	Student 1	Ja	Ja	Nein	Nein	Nein	Nein	Ja	Nein	Nein	
1	Student 2	Ja	Nein	Ja	Ja	Ja	Nein	Nein	Nein	Ja	
2	Student 3	Ja	Nein	Nein	Nein	Nein	Nein	Nein	Nein	Nein	
3	Student 4	Ja	Nein	Ja	Ja	Ja	Nein	Nein	Nein	Nein	
4	Student 5	Ja	Nein	Ja	Ja	Ja	Ja	Ja	Ja	Nein	

```
In [58]: df8 = df8.replace(["Ja"],[1])
df8 = df8.replace(["Nein"],[0])
df8.head()
```

Out[58]:

	Studenten ID	Analoger Spickzettel	Manipulierte Prüfungsmaterialien	Einblendung von Inhalten auf den Hauptbildschirm ohne Wissen des Prüfers	Einblendung von Inhalten auf dem zweiten Bildschirm ohne Wissen des Prüfers	Einblendung von Inhalten auf einem anderen Gerät	Virtuelle Kamera	Audio Signal im Ohr	Vortäuschung technischer Probleme	Vorbereitete Texte werden abgelesen	Übersetzungsprogra
0	Student 1	1	1	0	0	0	0	1	0	0	
1	Student 2	1	0	1	1	1	0	0	0	1	
2	Student 3	1	0	0	0	0	0	0	0	0	
3	Student 4	1	0	1	1	1	0	0	0	0	
4	Student 5	1	0	1	1	1	1	1	1	0	

```
In [59]: filtered_df8 = df8.loc[:,df8.columns!="Studenten ID"];
filtered_df8.sum(numeric_only=True)
```

```
Out[59]: Analoge Spickzettel                210
Manipulierte Prüfungsmaterialien          61
Einblendung von Inhalten auf den Hauptbildschirm ohne Wissen des Prüfers    83
Einblendung von Inhalten auf dem zweiten Bildschirm ohne Wissen des Prüfers  108
Einblendung von Inhalten auf einem anderen Gerät                          121
Virtuelle Kamera                        31
Audio Signal im Ohr                      55
Vortäuschung technischer Probleme        79
Vorbereitete Texte werden abgelesen     119
Übersetzungsprogramme                    95
Kommunikation mit anderen Kommilitonen während der Prüfung                166
Lösungen von Dritten werden komplett kopiert                               84
Prüfung wird von jemand anderen geschrieben                                57
dtype: int64
```

```
In [60]: df9 = pd.read_excel(r"C:\Users\stefa\Desktop\BA_Umfrage.xlsx", sheet_name='Spickzettel')
df9.head()
```

Out[60]:

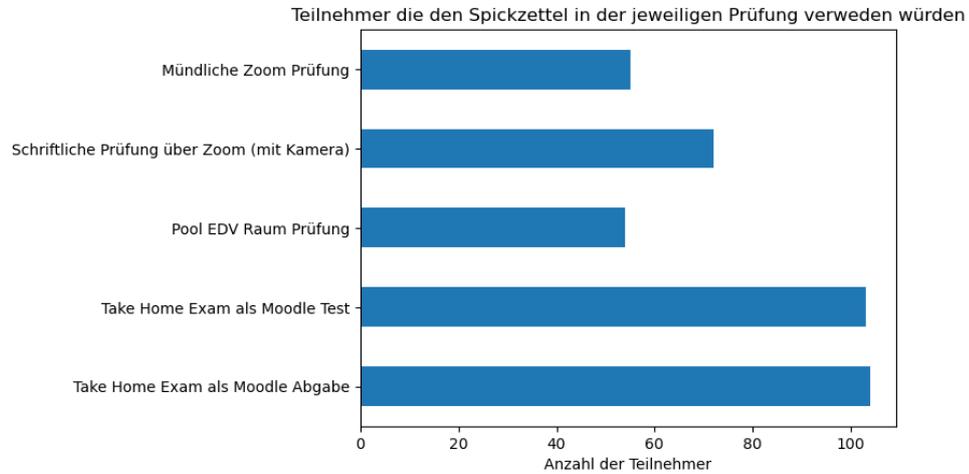
	Take Home Exam als Moodle Abgabe	Take Home Exam als Moodle Test	Pool EDV Raum Prüfung	Schriftliche Prüfung über Zoom (mit Kamera)	Mündliche Zoom Prüfung
0	NaN	NaN	NaN	1.0	NaN
1	1.0	1.0	1.0	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	1.0	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	1.0

```
In [61]: df9 = df9.fillna(0)
```

```
In [62]: df9.sum(numeric_only=True)
```

```
Out[62]: Take Home Exam als Moodle Abgabe      104.0  
Take Home Exam als Moodle Test      103.0  
Pool EDV Raum Prüfung      54.0  
Schriftliche Prüfung über Zoom (mit Kamera)  72.0  
Mündliche Zoom Prüfung      55.0  
dtype: float64
```

```
In [63]: df9.sum(numeric_only=True).plot.barh();  
plt.title("Teilnehmer die den Spickzettel in der jeweiligen Prüfung verwenden würden");  
plt.xlabel("Anzahl der Teilnehmer");
```



```
In [64]: df10 = pd.read_excel(r"C:\Users\stefa\Desktop\BA_Umfrage.xlsx", sheet_name='Kommunikation')  
df10.head()
```

```
Out[64]:
```

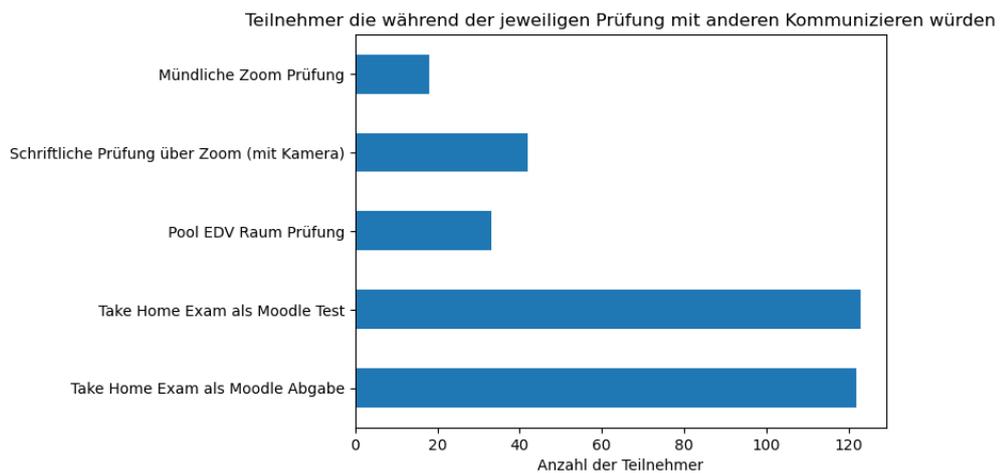
	Take Home Exam als Moodle Abgabe	Take Home Exam als Moodle Test	Pool EDV Raum Prüfung	Schriftliche Prüfung über Zoom (mit Kamera)	Mündliche Zoom Prüfung
0	1.0	1.0	NaN	1.0	NaN
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	1.0
4	1.0	1.0	NaN	1.0	NaN

```
In [65]: df10 = df10.fillna(0)
```

```
In [66]: df10.sum(numeric_only=True)
```

```
Out[66]: Take Home Exam als Moodle Abgabe      122.0  
Take Home Exam als Moodle Test      123.0  
Pool EDV Raum Prüfung      33.0  
Schriftliche Prüfung über Zoom (mit Kamera)  42.0  
Mündliche Zoom Prüfung      18.0  
dtype: float64
```

```
In [67]: df10.sum(numeric_only=True).plot.barh();  
plt.title("Teilnehmer die während der jeweiligen Prüfung mit anderen Kommunizieren würden");  
plt.xlabel("Anzahl der Teilnehmer");
```



#### 4. Auszug aus der Exceldatei für das Clustering

	A	B	C	D	E	F	G	H	I
1	<b>Schummelhaltung</b>	<b>Schummeln_TH_Abgab</b>	<b>Schummeln_TH_Tes</b>	<b>Schummeln_Pool_Prüfer</b>	<b>Schummeln_Schriftliche_Zoo</b>	<b>Schummeln_Mündliche_Zoo</b>	<b>Studiengang</b>		<b>Fachsemester</b>
2	Neutrale Haltung	Ja	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		3
3	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
4	Verwerflich wenn es andere machen	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
5	Wenn es meine Freunde machen ist es ok	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
6	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
7	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
8	Ich finde schummeln ok	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
9	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
10	Neutrale Haltung	Nein	Ja	Nein	Nein	Ja	Wirtschaftsingenieurwesen		1
11	Neutrale Haltung	Ja	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
12	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
13	Neutrale Haltung	Ja	Ja	Nein	Ja	Nein	Wirtschaftsingenieurwesen		1
14	Verwerflich wenn es andere machen	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		2
15	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
16	Verwerflich wenn es andere machen	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
17	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
18	Ich finde schummeln ok	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
19	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
20	Ich finde schummeln ok	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
21	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
22	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		1
23	Neutrale Haltung	Ja	Ja	Ja	Nein	Nein	Informationsmanagement im Gesundheitswesen		1
24	Ich finde schummeln ok	Nein	Nein	Nein	Nein	Nein	Betriebswirtschaft		4
25	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Betriebswirtschaft im Gesundheitswesen		9
26	Verwerflich wenn es andere machen	Nein	Nein	Nein	Ja	Nein	Sonstiges		3
27	Ich finde schummeln ok	Ja	Ja	Nein	Nein	Nein	Informationsmanagement und Unternehmenskommunikation		4
28	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Informationsmanagement und Unternehmenskommunikation		7
29	Verwerflich wenn es andere machen	Nein	Nein	Nein	Nein	Nein	Sonstiges		2
30	Neutrale Haltung	Nein	Ja	Nein	Nein	Nein	Wirtschaftspsychologie		6
31	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Sonstiges		1
32	Neutrale Haltung	Ja	Ja	Ja	Nein	Ja	Informationsmanagement und Unternehmenskommunikation		3
33	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein	Wirtschaftsingenieurwesen		8
34	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein			1

## 5. Clusterverfahren kMeans Code

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
sns.set()
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import LabelEncoder
from sklearn.decomposition import PCA

labelencoder = LabelEncoder()
encoder = OneHotEncoder(handle_unknown='ignore')
```

### Vorbereitung der Daten

```
In [2]: cluster = pd.read_excel(r"C:\Users\stefa\Desktop\Clustering.xlsx", sheet_name='Cluster')
cluster.head()
```

Out[2]:

	Schummelhaltung	Schummeln_TH_Abgabe	Schummeln_TH_Test	Schummeln_Pool_Prüfung	Schummeln_Schriftliche_Zoom	Schummeln_Mündliche_Zoom
0	Neutrale Haltung	Ja	Nein	Nein	Nein	Nein
1	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein
2	Verwerflich wenn es andere machen	Nein	Nein	Nein	Nein	Nein
3	Wenn es meine Freunde machen ist es ok	Nein	Nein	Nein	Nein	Nein
4	Neutrale Haltung	Nein	Nein	Nein	Nein	Nein

5 rows x 36 columns

```
In [3]: cluster = cluster.replace(["Ja"], [1])
cluster = cluster.replace(["Nein"], [0])
cluster = cluster.fillna("na")
cluster["Alter"] = cluster["Alter"].replace(["na"], 0)
cluster["Fachsemester"] = cluster["Fachsemester"].replace(["na"], 0)
cluster.head()
```

Out[3]:

	Schummelhaltung	Schummeln_TH_Abgabe	Schummeln_TH_Test	Schummeln_Pool_Prüfung	Schummeln_Schriftliche_Zoom	Schummeln_Mündliche_Zoom
0	Neutrale Haltung	1	0	0	0	0
1	Neutrale Haltung	0	0	0	0	0
2	Verwerflich wenn es andere machen	0	0	0	0	0
3	Wenn es meine Freunde machen ist es ok	0	0	0	0	0
4	Neutrale Haltung	0	0	0	0	0

5 rows x 36 columns

```
In [4]: one_hot_encoding_columns = ['Schummelhaltung', 'Studiengang', 'Geschlecht', 'Wohnort', 'Zufriedenheit_Studium',
                                   'Zufriedenheit_Hochschule', 'Motivation', 'Leistungsdruck']

for x in one_hot_encoding_columns :
    cluster[x] = labelencoder.fit_transform(cluster[x])
    encoder_cluster = pd.DataFrame(encoder.fit_transform(cluster[[x]]).toarray())

    classes = labelencoder.classes_
    desired_classes = ["Stimme voll und ganz zu", "Stimme zu", "Weder noch", "Stimme nicht zu",
                      "Stimme überhaupt nicht zu"]
    class_names = []
    for class_ in classes:
        if class_ in desired_classes:
            class_names.append(x + "_" + class_.replace(" ", "_"))
        else:
            class_names.append(class_)

    encoder_cluster.columns = class_names
    cluster = cluster.join(encoder_cluster)

columns_to_drop = ["Sonstiges", "na"]
for column in columns_to_drop:
    if column in cluster.columns:
        cluster.drop(column, axis=1, inplace=True)

cluster.drop([x], axis=1, inplace = True)

cluster.head()
```

Out[4]:

	Schummeln_TH_Abgabe	Schummeln_TH_Test	Schummeln_Pool_Prüfung	Schummeln_Schriftliche_Zoom	Schummeln_Mündliche_Zoom	Fachsemester	Not
0	1	0	0	0	0		3.0
1	0	0	0	0	0		1.0
2	0	0	0	0	0		1.0
3	0	0	0	0	0		1.0
4	0	0	0	0	0		1.0

5 rows x 72 columns

```
In [5]: cluster.to_excel("cluster_OneHotEncoder.xlsx", index=False)
```

```
In [6]: x1 = cluster.loc[:, :].values
x_stand1 = StandardScaler().fit_transform(x1)
X_stand1 = pd.DataFrame(data = x_stand1, columns = cluster.columns)
X_stand1.head()
```

Out[6]:

	Schummeln_TH_Abgabe	Schummeln_TH_Test	Schummeln_Pool_Prüfung	Schummeln_Schriftliche_Zoom	Schummeln_Mündliche_Zoom	Fachsemester	Not
0	1.811997	-0.563869	-0.21146	-0.274434	-0.305969		-0.389471
1	-0.551877	-0.563869	-0.21146	-0.274434	-0.305969		-1.180730
2	-0.551877	-0.563869	-0.21146	-0.274434	-0.305969		-1.180730
3	-0.551877	-0.563869	-0.21146	-0.274434	-0.305969		-1.180730
4	-0.551877	-0.563869	-0.21146	-0.274434	-0.305969		-1.180730

5 rows x 72 columns

```
In [7]: X_stand1.describe().T
```

Out[7]:

	count	mean	std	min	25%	50%	75%	max
Schummeln_TH_Abgabe	257.0	4.665529e-17	1.001951	-0.551877	-0.551877	-0.551877	-0.551877	1.811997
Schummeln_TH_Test	257.0	-6.652698e-17	1.001951	-0.563869	-0.563869	-0.563869	-0.563869	1.773460
Schummeln_Pool_Prüfung	257.0	1.784133e-16	1.001951	-0.211460	-0.211460	-0.211460	-0.211460	4.729021
Schummeln_Schriftliche_Zoom	257.0	6.911894e-17	1.001951	-0.274434	-0.274434	-0.274434	-0.274434	3.643869
Schummeln_Mündliche_Zoom	257.0	-1.092943e-16	1.001951	-0.305969	-0.305969	-0.305969	-0.305969	3.268305
...	...	...	...	...	...	...	...	...
Leistungsdruck_Stimme_nicht_zu	257.0	3.844741e-17	1.001951	-0.370362	-0.370362	-0.370362	-0.370362	2.700060
Leistungsdruck_Stimme_voll_und_ganz_zu	257.0	9.417456e-17	1.001951	-0.539868	-0.539868	-0.539868	-0.539868	1.852305
Leistungsdruck_Stimme_zu	257.0	6.393502e-17	1.001951	-0.765780	-0.765780	-0.765780	1.305857	1.305857
Leistungsdruck_Stimme_überhaupt_nicht_zu	257.0	1.079983e-17	1.001951	-0.290496	-0.290496	-0.290496	-0.290496	3.442383
Leistungsdruck_Weder_noch	257.0	9.719851e-17	1.001951	-0.491473	-0.491473	-0.491473	-0.491473	2.034699

72 rows x 8 columns

## Durchführung einer PCA

```
In [8]: from sklearn.decomposition import PCA
pca_1 = PCA(n_components=2)
principalComponents = pca_1.fit_transform(X_stand1)

principalDf1 = pd.DataFrame(data = principalComponents, columns = ["principal component 1", "principal component 2"])
principalDf1.head()
```

```
Out[8]:
```

	principal component 1	principal component 2
0	-0.081295	-0.393206
1	2.175168	-0.477266
2	4.449978	-1.913057
3	5.578461	-2.769866
4	2.076760	-1.034383

## Durchführung des kMeans

```
In [9]: from sklearn.cluster import KMeans

# KMeans-Modell erstellen
km_model = KMeans(n_clusters=3, random_state=0).fit(principalDf1)
```

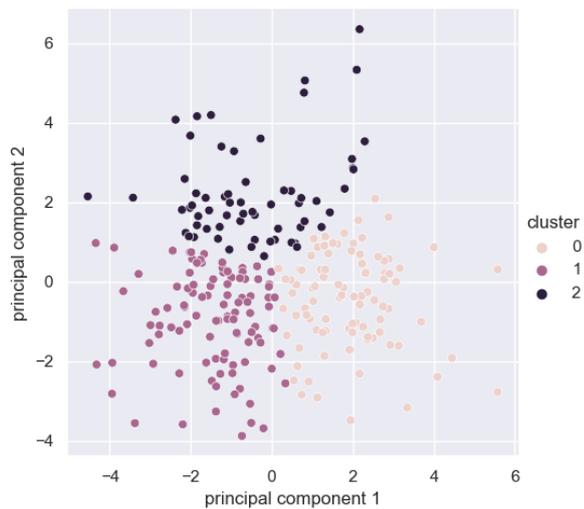
```
In [10]: km_model.labels_
```

```
Out[10]: array([[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1,
2, 2, 2, 2, 1, 0, 1, 1, 2, 0, 1, 1, 1, 1, 1, 2, 0, 2, 0, 1, 2, 2,
1, 0, 2, 1, 1, 2, 1, 2, 1, 0, 0, 2, 2, 0, 0, 1, 1, 1, 1, 1, 0, 0,
1, 0, 0, 2, 1, 1, 0, 0, 1, 2, 2, 1, 0, 2, 2, 2, 0, 1, 0, 0, 1, 1,
1, 0, 2, 1, 0, 0, 0, 2, 1, 1, 0, 0, 1, 0, 1, 1, 1, 2, 1, 2, 0, 2,
1, 1, 2, 2, 1, 2, 2, 1, 2, 2, 0, 2, 0, 0, 2, 2, 1, 1, 1, 0, 0, 0,
1, 1, 0, 1, 0, 1, 2, 0, 0, 1, 1, 0, 1, 2, 0, 0, 0, 2, 2, 1, 1, 1,
0, 0, 0, 2, 1, 0, 1, 1, 1, 0, 0, 2, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0,
1, 2, 0, 2, 2, 0, 1, 0, 2, 2, 2, 0, 0, 2, 0, 2, 1, 2, 1, 2, 1, 2,
1, 0, 1, 1, 1, 1, 0, 1, 2, 1, 2, 0, 2, 0, 1, 1, 0, 1, 1, 1, 1, 1,
0, 2, 0, 1, 1, 0, 0, 0, 1, 0, 1, 2, 1, 2, 1, 1, 0, 1, 2, 0, 1, 1,
2, 0, 1, 1, 0, 0, 2, 1, 2, 2, 0, 0, 1, 1, 1])
```

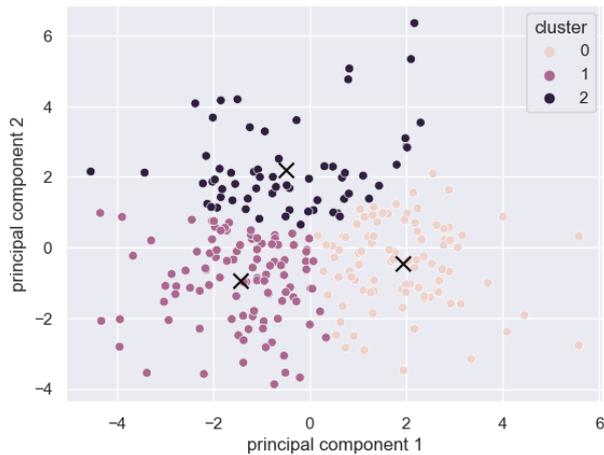
```
In [11]: km_model.cluster_centers_
```

```
Out[11]: array([[ 1.91893459, -0.46465394],
[-1.43169929, -0.94395237],
[-0.48426435,  2.2068461 ]])
```

```
In [12]: cluster_y = principalDf1.copy()
cluster_y["cluster"] = km_model.labels_
sns.relplot(x = "principal component 1", y = "principal component 2", hue = "cluster", data = cluster_y)
plt.show()
```



```
In [13]: ▶ centers = km_model.cluster_centers_
g = sns.scatterplot(data=cluster_y, x="principal component 1", y="principal component 2", hue="cluster")
g.scatter(centers[:, 0], centers[:, 1], marker='x', s=100, color='#050505')
plt.show()
```



```
In [14]: ▶ cluster_y = cluster.copy()
cluster_y["cluster"] = km_model.labels_
cluster_y.head()
```

Out[14]:

	Schummeln_TH_Abgabe	Schummeln_TH_Test	Schummeln_Pool_Prüfung	Schummeln_Schriftliche_Zoom	Schummeln_Mündliche_Zoom	Fachsemester	Not
0	1	0	0	0	0	0	3.0
1	0	0	0	0	0	0	1.0
2	0	0	0	0	0	0	1.0
3	0	0	0	0	0	0	1.0
4	0	0	0	0	0	0	1.0

5 rows x 73 columns

```
In [15]: ▶ cluster_y.to_excel("cluster_kMeans.xlsx", index=False)
```

```
In [16]: ▶ mask0 = cluster_y["cluster"] == 0
dfcluster0 = cluster_y[mask0]
dfcluster0.head()
```

Out[16]:

	Schummeln_TH_Abgabe	Schummeln_TH_Test	Schummeln_Pool_Prüfung	Schummeln_Schriftliche_Zoom	Schummeln_Mündliche_Zoom	Fachsemester	Not
1	0	0	0	0	0	0	1.0
2	0	0	0	0	0	0	1.0
3	0	0	0	0	0	0	1.0
4	0	0	0	0	0	0	1.0
5	0	0	0	0	0	0	1.0

5 rows x 73 columns

```
In [17]: ▶ mask1 = cluster_y["cluster"] == 1
dfcluster1 = cluster_y[mask1]
dfcluster1.head()
```

Out[17]:

	Schummeln_TH_Abgabe	Schummeln_TH_Test	Schummeln_Pool_Prüfung	Schummeln_Schriftliche_Zoom	Schummeln_Mündliche_Zoom	Fachsemester	Not
0	1	0	0	0	0	0	3.0
19	0	0	0	0	0	0	1.0
21	1	1	1	0	0	0	1.0
26	0	0	0	0	0	0	7.0
28	0	1	0	0	0	0	6.0

5 rows x 73 columns

```
In [18]: mask2 = cluster_y["cluster"] == 2
dfcluster2 = cluster_y[mask2]
dfcluster2.head()
```

Out[18]:

	Schummeln_TH_Abgabe	Schummeln_TH_Test	Schummeln_Pool_Prüfung	Schummeln_Schriftliche_Zoom	Schummeln_Mündliche_Zoom	Fachsemester	Nr
11	1	1	0	1	0	1.0	
22	0	0	0	0	0	4.0	
23	0	0	0	0	0	9.0	
24	0	0	0	1	0	3.0	
25	1	1	0	0	0	4.0	

5 rows x 73 columns

```
In [19]: result0 = dfcluster0.describe().T
result1 = dfcluster1.describe().T
result2 = dfcluster2.describe().T
```

```
In [20]: with pd.ExcelWriter('cluster_describe.xlsx') as writer:
result0.to_excel(writer, sheet_name='cluster 0')
result1.to_excel(writer, sheet_name='cluster 1')
result2.to_excel(writer, sheet_name='cluster 2')
```

## Auswertung der erstellten Cluster

```
In [21]: demo = cluster_y[["Notendurchschnitt", "Fachsemester", "Alter", "cluster"]]
Hobbie = cluster_y[["Fußball", "Basketball", "Tennis", "Tanzen", "Yoga", "Schwimmen", "Fitnessstudio",
"Pilates", "Klettern", "Kampfsport", "Reiten", "Handball", "Joggen/Laufen", "Schach", "Poker",
"Freunde Treffen", "Lesen", "Malen", "Kino", "Bar/Club", "cluster"]]
Studiengang = cluster_y[["Betriebswirtschaft", "Betriebswirtschaft im Gesundheitswesen",
"Data Science Management", "Digital Enterprise Management",
"Digitales Medizin- und Pflegemanagement", "Game-Production und Management",
"Information Management Automotive", "Informationsmanagement im Gesundheitswesen",
"Informationsmanagement und Unternehmenskommunikation", "Physician Assistant",
"Wirtschaftsinformatik", "Wirtschaftsingenieurwesen", "Wirtschaftspsychologie", "cluster"]]
Geschlecht = cluster_y[["Weiblich", "männlich", "nicht angegeben", "cluster"]]
Wohnort = cluster_y[["Bei deinen Eltern", "Eigene Wohnung", "Studentenwohnheim", "WG-Zimmer", "cluster"]]
Studium_Zufrieden = cluster_y[["Zufriedenheit Studium Stimme nicht zu", "Zufriedenheit Studium Stimme voll und ganz zu",
"Zufriedenheit Studium Stimme zu", "Zufriedenheit Studium Stimme überhaupt nicht zu",
"Zufriedenheit Studium Weder noch", "cluster"]]
Hochschule_Zufrieden = cluster_y[["Zufriedenheit Hochschule Stimme nicht zu",
"Zufriedenheit Hochschule Stimme voll und ganz zu",
"Zufriedenheit Hochschule Stimme zu",
"Zufriedenheit Hochschule Stimme überhaupt nicht zu",
"Zufriedenheit Hochschule Weder noch", "cluster"]]
Motivation = cluster_y[["Motiviert", "Sehr motiviert", "Sehr unmotiviert", "Unmotiviert", "Motivation Weder noch",
"cluster"]]
Leistungsdruck = cluster_y[["Leistungsdruck Stimme nicht zu", "Leistungsdruck Stimme voll und ganz zu",
"Leistungsdruck Stimme zu", "Leistungsdruck Stimme überhaupt nicht zu",
"Leistungsdruck Weder noch", "cluster"]]
Schummeln = cluster_y[["Schummeln_TH_Abgabe", "Schummeln_TH_Test", "Schummeln_Pool_Prüfung",
"Schummeln_Schriftliche_Zoom", "Schummeln_Mündliche_Zoom", "Ich finde schummeln ok",
"Neutrale Haltung", "Verwerflich wenn es andere machen",
"Wenn es meine Freunde machen ist es ok", "cluster"]]
```

```
In [22]: demo.describe()
```

Out[22]:

	Notendurchschnitt	Fachsemester	Alter	cluster
count	257.000000	257.000000	257.000000	257.000000
mean	2.141753	3.984436	21.599222	0.887160
std	0.485037	2.532552	7.675987	0.769861
min	1.250000	0.000000	0.000000	0.000000
25%	1.750000	2.000000	20.000000	0.000000
50%	2.141753	4.000000	22.000000	1.000000
75%	2.250000	6.000000	24.000000	1.000000
max	3.750000	11.000000	56.000000	2.000000

```
In [23]: grouped = demo.groupby(["cluster"])
for name, group in grouped:
    print(name)
    print(group.describe())
```

```
0
  Notendurchschnitt  Fachsemester    Alter  cluster
count      92.000000    92.000000  92.000000    92.0
mean         2.153743     3.336957  20.608696     0.0
std          0.483358     2.326518   8.083298     0.0
min          1.250000     1.000000   0.000000     0.0
25%          2.141753     1.000000  20.000000     0.0
50%          2.141753     3.000000  21.000000     0.0
75%          2.250000     5.000000  23.250000     0.0
max          3.750000    11.000000  46.000000     0.0

1
  Notendurchschnitt  Fachsemester    Alter  cluster
count     102.000000   102.000000  102.000000   102.0
mean         2.061704     3.764706  21.068627     1.0
std          0.486720     2.510084   8.032434     0.0
min          1.250000     0.000000   0.000000     1.0
25%          1.750000     1.000000  20.000000     1.0
50%          2.141753     3.000000  22.000000     1.0
75%          2.250000     6.000000  24.000000     1.0
max          3.250000    11.000000  56.000000     1.0

2
  Notendurchschnitt  Fachsemester    Alter  cluster
count      63.000000    63.000000   63.000000    63.0
mean         2.253846     5.285714  23.904762     2.0
std          0.468028     2.419201   5.912573     0.0
min          1.250000     1.000000   0.000000     2.0
25%          1.750000     3.000000  22.000000     2.0
50%          2.250000     6.000000  23.000000     2.0
75%          2.750000     7.000000  24.000000     2.0
max          3.250000    11.000000  43.000000     2.0
```

```
In [24]: data_frames = [Hobbie, Studiengang, Geschlecht, Wohnort, Studium_Zufrieden,
Hochschule_Zufrieden, Motivation, Leistungsdruck, SchummeIn]
```

```
for i, data in enumerate(data_frames):
    grouped = data.groupby(data.columns[-1])
    fig, ax = plt.subplots(figsize=(15,7))
    for key, group in grouped:
        mean_group = group.mean().drop(group.columns[-1])
        mean_group.plot(ax=ax, kind='line', label=key)
        ax.set_xlabel("", fontsize=10)

    # Berechnung des Mittelwerts aller Daten
    mean_all = data.mean().drop(data.columns[-1])
    # Hinzufügen der Linie zum Diagramm
    mean_all.plot(ax=ax, kind='line', label='Mittelwert aller Daten', color='black')

    ax.set_xticks(range(len(mean_group.index)))
    ax.set_xticklabels(mean_group.index, rotation=90, fontsize=8)
    ax.set_ylabel('Mittelwert', fontsize=10)
    plt.legend()
    plt.show()
```

