Unterrichtsreihe „Datenexploration – Datendetektiv:innen bei der Arbeit“

## Kernidee

In dieser Unterrichtseinheit geht es um Data Science Inhalte für die 8.-10. Klasse. Anhand eines fiktiven Rahmenbeispiels einer Online-Plattform, die Werbung für Jugendliche passgenau schalten möchte, werden Schülerinnen und Schüler motiviert, als Datendetektiv:innen in einem Datensatz nach Spuren und Mustern zu suchen, um so die Online-Plattform zu beraten.

Es stehen **Arbeitsblätter**, **PowerPoint-Präsentationen**, **Anleitungen**, der **YOU-PB-Datensatz**, eine **Variablenliste** und ein **Überblick** über die einzelnen Stunden zur Verfügung. Für die Datenanalyse wird die Software **CODAP** ([codap.concord.org](https://codap.concord.org/app/static/dg/de/cert/index.html)) genutzt, die im Internet frei und kostenlos zugänglich ist.

Gearbeitet wird mit Daten von über 1000 Schülerinnen und Schülern, die zu vielen Merkmalen im Freizeit- und Medienbereich Angaben gemacht haben (YOU-PB steht für Mediennutzung Jugendlicher im Raum Paderborn). Der Datensatz liegt für diese Unterrichtsreihe in verschiedenenLevor. Zum einen steht eine reduzierte Version mit 50 Variablen zur Verfügung. Zum anderen kann die Vollversion mit über 160 Variablen im Unterricht eingesetzt werden. Es wird empfohlen, aufgrund der besseren Übersichtlichkeit den Datensatz mit der reduzierten Variablenanzahl zu verwenden. Hier sind multivariate und interessante Entdeckungen möglich! Die Lehrkraft hat jedoch die Möglichkeit je nach Leistungsfähigkeit der Klasse/des Kurses zu differenzieren, oder für besonders leistungsfähige Schülerinnen und Schüler ein binnendifferenziertes Angebot zu unterbreiten (siehe unten).

Der Link zum Datensatz, mit dem im Projekt gearbeitet wird (50 Variablen):

<https://tinyurl.com/you-pb50>

Die Unterrichtsreihe umfasst ca. 8 Unterrichtsstunden. In den ersten Stunden werden die Schülerinnen und Schüler in die Daten und das Arbeiten mit CODAP eingeführt. Das Herzstück der Unterrichtsreihe ist eine eigenständige Data Science Projektarbeit der Schülerinnen und Schüler in den Stunden 5 und 6 mit Präsentationen in Stunde 7. Die Projektarbeit findet in Kleingruppen statt, in denen sich die Schülerinnen und Schüler eigenständig als Data-Science-Expert:innen mit dem Datensatz beschäftigen. Dafür ordnen sie sich vier inhaltlichen Gruppen der YOU-PB-Daten zu, analog zu den Kundenwünschen der Online-Plattform (diese Zuordnung geschieht in Stunde 4):

* Kunde 1 möchte TikTok bewerben,
* Kunde 2 möchte LetsPlay\_YouTube Videos bewerben,
* Kunde 3 möchte Online-Zeitungen bewerben,
* Kunde 4 möchte feste Spielekonsolen bewerben.

In der letzten Stunde findet eine Reflektion über das Vorgehen für die Datenexploration statt und persönliche und gesellschaftliche Aspekte können diskutiert werden. Außerdem kann hier als Exkurs das Thema Datenbereinigung angesprochen und in CODAP bearbeitet werden, die in der Arbeit von echten Data-Scientists viel Zeit beansprucht.

### Zielgruppe

Mathematik ab Klasse 8 (alle Schulformen)

Informatik ab Klasse 8 (alle Schulformen)

### Vorkenntnisse

Prozentbegriff

### Zeitlicher Umfang

8 bis 10 Unterrichtsstunden a 45 Minuten

### Ziele

Statistische Lernziele

* Lernende explorieren multivariate Daten selbständig
* Lernende führen Gruppenvergleiche durch
* Lernende nutzen Zeilen-, Spalten- und Zellenprozente und interpretieren die Ergebnisse

Informatik Lernziele

* Lernende interpretieren Daten als Information
* Lernende erhalten ein Verständnis für den Aufbau und die Funktionsweise von Informatiksystemen
* Lernende nutzen Datenbanksoftware
* Lernende diskutieren gesellschaftliche Verantwortung und Auswirkung des Nutzens von Daten

Medienziele

* Lernende nutzen digitale Werkzeuge zum Analysieren von Daten
* Lernende erstellen Präsentationen und üben Präsentieren

Überblick über die Unterrichtsreihe

|  |  |  |  |
| --- | --- | --- | --- |
| **Stunde** | **Thema** | **Ziele** | **Material** |
| 1 | **Einführung**  In dieser Stunde wird in das Projekt „Datendetektiv:innen bei der Arbeit” eingeführt. Dazu wird der komplette Rahmen der Unterrichtsreihe durch die Lehrkraft über **Präsentation 1** aufgezeigt. In zwei Arbeitsphasen werden die SuS an die vorliegenden Umfragedaten und an die Datenanalyse in CODAP herangeführt.  Das Dokument **Stunde\_1\_Überblick** bietet Gestaltungshinweise für diese Stunde.  Mit **Arbeitsblatt 1** erkunden SuS den Datensatz selbständig.  Didaktische Hinweise  Die Einführung in CODAP kann anhand eines Videos oder durch Demonstration durch die Lehrkraft geschehen (**siehe Stunde 1 Überblick**)  Das Dokument **Hinweise\_Lehrkraft** enthält einen Überblick über das Analysieren kategorialer Merkmale mit CODAP und möglicher Schüler:innenschwierigkeiten.  Ergänzungsmöglichkeit  Bei APPCamps gibt es Lernmaterial zum Hinterlassen von Daten, was als Einstieg genutzt werden kann, um „Datenspuren“ zu thematisieren: <https://drive.google.com/file/d/1eb0_qlnKP-63H8El0OIsDUg3xXgDtVUA/view> | SuS explorieren selbständig die YOU-PB Daten mit CODAP | Stunde\_1\_Überblick  Präsentation\_1  Arbeitsblatt\_1  Arbeitsblatt\_1\_Lösungsvorschlag  Variablenliste\_YOU-PB\_2021\_50  Kurzeinführung in CODAP:  Anleitung\_CODAP  ODER  [https://YOU-PBtu.be/2z5H4anfhWM](https://youtu.be/2z5H4anfhWM)  (ca. 5 min)  Hinweise\_Lehrkraft |
| 2-3 | In diesen beiden Stunden geht es darum, dass die Schülerinnen und Schüler erste Erfahrungen als Data-Science- Expertinnen und Experten erlangen, um anschließend selbständig arbeiten zu können. Besonders das Arbeiten mit kategorialen Daten liegt hier im Fokus.  Zunächst kann, sofern noch nicht geschehen, Arbeitsblatt 1 besprochen werden.  **Einführung in nötige Grundbegriffe und Erwartungshaltung aufbauen**  Die SuS erhalten Informationen über grundlegende statistische Fachbegriffe. Die kennengelernten, statistischen Grundbegriffe übertragen die SuS anhand von **Arbeitsblatt 2** auf den vorliegenden Datensatz.  **CODAP und Analysemethoden**  Es werden Auswertungsmöglichkeiten bezüglich verschiedener Prozente (Zeilenprozente, Spaltenprozente und Zellenprozente, siehe **Hinweise Lehrkraft**) und damit verbundener Aussagen erarbeitet.  Zunächst steht thematisch das „Modellieren der Daten“ an. Interessierende Merkmale werden umcodiert, so dass die Ausprägungen von z. B. sieben (täglich, mehrmals pro Woche, … nie) auf zwei (z. B. häufig, selten) reduziert werden. Dies vereinfacht die Auswertungen.  Die Arbeitsblätter in der Übersicht  Es gibt **Anleitungen**, anhand derer Schüler:innen sich Inhalte selbst erarbeiten können. Alternativ sollten die Inhalte der Anleitungen gemeinsam mit der Lehrkraft erarbeitet werden. Weiterhin gibt es **Arbeitsblätter**, die im Hinblick auf die spätere Projektarbeit bearbeitet und im Plenum besprochen werden sollten. Bei den Arbeitsblättern werden verschiedene Verteilungen miteinander verglichen werden. Die Arbeitsblätter sind gestuft aufgebaut. Wichtig ist, dass die gefundenen Zusammenhänge nur für die Personen der Stichprobe gelten und so einfach verallgemeinert werden können, weil die Stichprobe (die Daten) nicht repräsentativ erhoben wurden.  Die Anleitungen:   * Anleitung 1: SuS lernen, Merkmale umzucodieren von sieben Ausprägungen zu zwei Ausprägungen * Anleitung 2: SuS lernen die verschiedenen Auswertungsmöglichkeiten: Zeilen-, Spalten- und Zellenprozente * Anleitung 3: SuS lernen, Daten auszublenden, um Analysen für Teilgruppen durchzuführen * Infoblatt Glossar: Hier werden wichtige Fachbegriffe kurz erklärt   Die Arbeitsblätter:   * Arbeitsblatt 3: zwei binäre Verteilungen (Verteilungen mit zwei Ausprägungen) werden miteinander verglichen * Arbeitsblatt 4: eine binäre Verteilung wird in Beziehung gesetzt mit einer Verteilung mit 7 Ausprägungen. Entweder wird hier umcodiert (empfohlen) oder präzise Aussagen müssen getätigt werden * Arbeitsblatt 5 (Fortgeschritten): zwei siebenstufige Verteilungen werden miteinander verglichen. Hier sollte umcodiert werden * Arbeitsblatt 6 (optional, z. B. Bonus):   Didaktische Hinweise  Die **Präsentation 2** zu Sitzung 2+3 kann als Lehrerhintergrundinformation oder zur Präsentation im Unterricht begleitend zu den Arbeitsblättern 3-6a genutzt werden.  Mehrere Unterrichtserprobungen haben gezeigt, dass das Umcodieren von Merkmalen (sieben verschiedene Ausprägungen werden auf zwei reduziert) ein sinnvolles Vorgehen ist und gleichzeitig das Modellieren thematisiert wird. Dies wird in Anleitung\_CODAP\_umcodieren (diese Anleitung ist im YouTube-Video umgesetzt) erklärt.  Wir schlagen vor, mit der Anleitung (Anleitung\_CODAP\_umcodieren) in die Exploration zu starten und dann die Anleitung\_CODAP\_2dimProzente weiter zu nutzen.  Eine Differenzierungsmöglichkeit besteht hier erneut darin, motivierte SuS Auswertungen mit den binären Merkmalen und zusätzlich mit den Merkmalen mit sieben Ausprägungen machen zu lassen (die Arbeitsblätter 3-5 lassen sich auf beide Arten bearbeiten).  Je nach Kurs/Klasse können sich die SuS die Auswertungsmöglichkeiten anhand der Anleitungen selbst erarbeiten und bei den Arbeitsblättern 3-5 anwenden. Alternativ kann die Lehrkraft mithilfe der PowerPoint in die Analysemethoden einführen und die SuS bearbeiten dann die Arbeitsblätter. Dies ist die notwendige Voraussetzung für die eigenständige Schülerexploration in den folgenden Stunden. *Erfahrungsgemäß benötigen SuS Unterstützen beim Bearbeiten und Interpretieren der prozentualen Auswertungen mit Zeilen-, Spalten- oder Zellenprozenten. Einen Einstieg auch zur Diskussion kann Arbeitsblatt 6 bieten.*  **Technik**  In dieser Stunde sollte den SuS auch gezeigt werden, wie sie Graphiken aus CODAP in eine Word-Datei oder eine PowerPoint Präsentation kopieren können. Ggf. ist auch das Weitergeben des CODAP-Dokuments per Link eine gute Möglichkeit, um Schüler:innenbearbeitungen zu dokumentieren, bzw. zu kontrollieren. | SuS können statistische Fachbegriffe auf die YOU-PB Daten anwenden  SuS manipulieren Daten durch Umcodieren der Ausprägungen  SuS stellen Zusammenhänge zwischen zwei kategorialen Variablen her und unterscheiden zwischen Zeilen-, Spalten- und Zellenprozenten | Präsentation 2  Arbeitsblatt 2  Infoblatt\_Glossar  Arbeitsblatt\_2  PowerPoint-Präsentation 2 (optional)  Anleitung 1  ([https://YOU-PBtu.be/qcK\_ZZsWfbQ](https://youtu.be/qcK_ZZsWfbQ))  Anleitung 2  Anleitung 3  Arbeitsblatt 3  Arbeitsblatt 4  Arbeitsblatt 5  Arbeitsblatt 6 (Bonus)  YouTube-Video zum Kopieren des Arbeitsstands per Link:  <https://youtu.be/otLuX8hhtq8> |
| 4 | **Erwartungshaltung aufbauen und passende Fragen stellen**  In dieser Stunde werden die Schülerinnen und Schüler in Kleingruppen für die komplette restliche Unterrichtseinheit eingeteilt und vier verschiedenen inhaltlichen Bereichen der Daten/Kunden der Online-Plattform zugeordnet (**Arbeitsblatt\_7**). Jede Gruppe sollte vier SuS umfassen, damit die spätere Think-pair-share Phase gut funktioniert.  Die Erfahrung hat gezeigt, dass das Stellen von geeigneten (statistischen) Fragestellungen für die Auswertung der Daten für SuS eine Hürde darstellt, deshalb liegt ein weiterer Fokus dieser Stunde darauf, die SuS anhand von **Arbeitsblatt 8** in Kleingruppen geeignete Fragen erarbeiten zu lassen. Hierfür wird die Methode Think-Pair-Share genutzt. Anhand von Arbeitsblatt 8 findet eine theoretische Auseinandersetzung mit den Daten, resp. der Variablenliste, statt, um eine Erwartungshaltung für die kommenden Stunden aufzubauen. Diese Erwartungshaltung ist zentral für die eigene Datenanalyse und soll auf Plakaten festgehalten werden. Die hierzu erstellten Plakate sollten im Verlauf der weiteren Unterrichtseinheit sichtbar für alle SuS im Klassenraum hängen.  Didaktische Hinweise  **Präsentation 3** fasst die Inhalte der letzten und dieser Stunde noch einmal zusammen.  **Lehrkrafthinweise 4** gibt Hinweise für diese Stunde. | SuS stellen geeignete statistische Fragen | Arbeitsblatt 7  Arbeitsblatt 8  Präsentation 3  Lehrkrafthinweise 4  Plakate |
| 5-6 | **Projektarbeit in Kleingruppen**  In diesen beiden Stunden wird das Vorgehen der Datenexploration von den Kleingruppen zunächst geplant (**Arbeitsblatt 9**). Anschließend findet die Datenexploration in den Kleingruppen statt (**Arbeitsblatt 10**) und die Präsentation der Ergebnisse wird vorbereitet (mit **Präsentation\_Vorlage**). Zu Beginn der sechsten Stunde können evtl. Kriterien für gute statistische Präsentationen besprochen werden (**Anleitung\_Hinweise\_Präsentation**), um die Erstellung der Präsentationen vorzubereiten. Die Lehrkraft steht in den Stunden 5-6 vor allem unterstützend für Fragen und Anregungen zur Verfügung, die SuS arbeiten möglichst selbständig mit CODAP und PowerPoint.  Didaktische Hinweise  Es hat sich in verschiedenen Erprobungen gezeigt, dass für dieser Projektphase auch zwei weitere Doppelstunden und ggf. Hausaufgaben aufgewendet werden können. | SuS explorieren die Daten nach eigenen Fragestellungen  SuS dokumentieren ihre Explorationen in PowerPoint  SuS bereiten eine Präsentation vor | Arbeitsblatt\_9  Arbeitsblatt\_10  Anleitung\_Hinweise\_Präsentation  Leere Präsentation\_Vorlage für Schüler:innenpräsentationen |
| 7 | **Präsentationen der Kleingruppen**  In dieser Stunde finden die Präsentationen der SuS-Gruppen statt. Dabei kann immer eine Gruppe einen speziellen Feedback-Auftrag bekommen, damit eine inhaltliche Diskussion angeregt wird (**Arbeitsblatt 11**). Anhand von **Arbeitsblatt 12**, das sich gut als Hausaufgabe und Ergebnissicherung eignet, können die SuS überprüfen, ob sie eine Datenanalyse durchführen können. | SuS präsentieren die Ergebnisse ihrer Projektarbeit  SuS geben Feedback zu anderen Präsentationen | PowerPoint-Präsentationen der SuS  Arbeitsblatt\_11  Arbeitsblatt\_12 |
| 8 | **Reflektieren**  In dieser Stunde findet eine gemeinsame Reflexion des gesamten Projekts statt. Die einzelnen Schritte der durchgeführten Datenanalyse werden den Stationen des PPDAC-Kreislaufs zugeordnet (**Präsentation 8**).  Weiter können unter anderem persönliche und gesellschaftliche Auswirkungen einer Datenexploration angesprochen werden und versucht werden, „über den Tellerrand” zu schauen. | SuS reflektieren den Datenanalyseprozess  SuS reflektieren gesellschaftliche Auswirkungen von Datenexplorationen | Präsentation 8 |
| Abschluss | **Zu Forschungszwecken** bitten wir darum, dass die SuS am Ende der Unterrichtsreihe eine anonyme Umfrage ausfüllen und Feedback geben, wie Ihnen die Unterrichtsreihe gefallen hat. Der Link zur Umfrage lautet: <https://umfrage-ddi.cs.uni-paderborn.de/limesurvey/index.php/545222?lang=de> |  | <https://umfrage-ddi.cs.uni-paderborn.de/limesurvey/index.php/545222?lang=de> |
| Exkurs  9 | **Datenbereinigung als Detektivarbeit**  Der Bereich Datenbereinigung kann als Exkurs in einer eigenen Stunde thematisiert werden. Hierzu gibt es die nebenstehende CODAP-Umgebung, in der sich der unbereinigte YOU-PB-Datensatz befindet. Dort findet sich ebenfalls ein Textfeld mit Erklärungen und ersten Schritten, wie mit Hilfe von CODAP eine Datenbereinigung vorgenommen werden kann. |  | Link zum unbereinigten Datensatz:  <https://tinyurl.com/you-pb50roh> |

# Weitere Informationen

Möglichkeit zur Differenzierung:

Die Unterrichtsreihe ist so angelegt, dass mit einem „kleinen“ Datensatz gearbeitet wird. Dieser enthält 50 Variable und bietet reichhaltig Möglichkeiten für Explorationen.

Für besonders motivierte Schülerinnen und Schüler kann stattdessen auch mit dem großen Datensatz gearbeitet werden, der alle 160 Variablen erhält, die in der Umfrage erhoben wurden. Dies erfordert jedoch ein hohes Engagement und gutes Arbeiten mit der Variablenliste seitens der Schülerinnen und Schüler!

Eine weitere Möglichkeit zur Differenzierung besteht darin, besonders motivierte Schülerinnen und Schüler mit den „normalen“ Merkmalen mit allen sieben Ausprägungen arbeiten zu lassen. Der Standardfall sollte das Arbeiten mit binären Merkmalen sein, also Merkmalen, die zuvor von den SuS umcodiert wurden, wie in Anleitung\_CODAP\_umcodieren in Stunde 2+3 beschrieben.

Zum Umgang mit den Aufgaben:

In der Unterrichtsreihe wird viel mit Arbeitsblättern gearbeitet. Um den Lernprozess zu dokumentieren und gleichzeitig die Motivation hoch zu halten, kann man die Aufgaben auch direkt in einer PowerPoint-Präsentation erarbeiten lassen. Neue Aufgaben können auf neuen Folien bearbeitet werden und am Ende können aus den verschiedenen Schulstunden Ergebnisse genommen werden, um die Abschlusspräsentation zu gestalten.

Teilnahme an der Umfrage:

Wer mit seiner Klasse an der Umfrage teilnehmen möchte, kann dies gerne machen. Die Daten werden vollständig anonym erfasst (eventuell mit der Schulleitung abklären). Die Daten fließen dann jährlich in eine Neuauflage des Datensatzes ein, der per CODAP-Link zugänglich bleibt.

Über diesen Link können SuS an der Umfrage teilnehmen (Achtung: 161 Fragen!, genügend Zeit einplanen): <http://go.upb.de/JIM-Umfrage>

# Tools + Daten

CODAP: Common Online Data Analysis Platform

CODAP (codap.concord.org) ist eine kostenlose, browserbasierte Software, die für den Einsatz ab Klasse 3 konzipiert ist. Sie unterstützt Lernende dabei, Statistik und Data Science zu lernen und zu verwenden. Um die Plattform zu benutzen, ist keine Registrierung nötig.

Im Toolkit dieser Website gibt es viel Lernmaterial zur Arbeit mit CODAP.

Der YouTube-Kanal von Susanne Podworny bietet weitere Lernvideos zum Umgang mit CODAP: <https://youtube.com/playlist?list=PLzhRG7IPqbqvKH5X6TIIfF8IzATgxHjQP&feature=shared>

CODAP ist in weiten Teilen auf Deutsch verfügbar. Sollte die Sprache nicht auf Deutsch eingestellt sein, so kann rechts oben die Sprache umgestellt werden.

Die YOU-PB Daten

Die hier verwendeten YOU-PB Daten entstammen einer Paderborner Replikationsstudie der offiziellen JIM-Studie, die diese Ergebnisse lokal nachzeichnet und zur Auswertung bereitstellt. Die JIM-Studie untersucht seit 1998 jährlich den Medienumgang von Zwölf- bis 19-Jährigen, um Trends und Entwicklungen im digitalen Verhalten der Jugend abzubilden und neue Impulse für Bildung und Kultur abzuleiten.

Die Daten stehen in CODAP in vier Varianten zur Verfügung:

Wir empfehlen Variante 1 für den Unterricht:

50 Variablen, binär codiert: <https://tinyurl.com/you-pb-50binaer> (Lang: <https://codap.concord.org/app/static/dg/de/cert/index.html#shared=https%3A%2F%2Fcfm-shared.concord.org%2F9IuhADJvA994rSs6MLct%2Ffile.json>)

50 Variablen, bis zu 7 Ausprägungen: <https://tinyurl.com/you-pb-50> (Lang: <https://codap.concord.org/app/static/dg/de/cert/index.html#shared=https%3A%2F%2Fcfm-shared.concord.org%2FH94YlIcxyFEYY9Ou8ZeL%2Ffile.json>)

160 Variablen, bis zu 7 Ausprägungen: <https://tinyurl.com/you-pb-160> (Lang: <https://codap.concord.org/app/static/dg/de/cert/index.html#shared=https%3A%2F%2Fcfm-shared.concord.org%2FDohqN1dGy7r2mf9iqL2j%2Ffile.json>)

160 Variablen, binär codiert: <https://tinyurl.com/you-pb-160binaer> (Lang: <https://codap.concord.org/app/static/dg/de/cert/index.html#shared=https%3A%2F%2Fcfm-shared.concord.org%2FstEcYRejbsfAPmTy9a26%2Ffile.json>)

Variablenliste für den 50er Datensatz: <https://www.prodabi.de/wp-content/uploads/Variablenliste_YOU-PB_2021_50.pdf>

Variablenliste für den 160er Datensatz: <https://www.prodabi.de/wp-content/uploads/Variablenliste_YOU_PB_161.pdf>