

Unterrichtsreihe „Datenexploration – Datendetektive bei der Arbeit“

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 Datendetektive in einem Datensatz nach Spuren und Mustern zu suchen, um so die Online-Plattform zu beraten.

Es stehen Arbeitsblätter, PowerPoint-Präsentationen, Anleitungen, der JIM-Datensatz, Variablenlisten und dieser Überblick über die einzelnen Stunden zur Verfügung. Für die Datenanalyse wird die Software CODAP (codap.concord.org) 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 (JIM-PB für Jugend-Information-Medien Paderborn). Der Datensatz liegt für diese Unterrichtsreihe in zwei Versionen vor. 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 Variablenliste 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/jim-pb50>

Die Unterrichtsreihe umfasst 8 Unterrichtsstunden. 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-Experten mit dem Datensatz beschäftigen. Dafür ordnen sie sich vier inhaltlichen Gruppen der JIM-Daten, analog zu den Kundenwünschen der Online-Plattform, zu (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 eines echten Data-Scientists viel Zeit beansprucht.

Überblick über die Unterrichtsreihe

Stunde	Thema	Material
1	Einführung In dieser Stunde wird in das Projekt „Datendetektive bei der Arbeit“ eingeführt. Dazu wird der komplette Rahmen der Unterrichtsreihe 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.	Stunde_1_Überblick Arbeitsblatt_1_Datensatz Variablenliste_JIM_2021_reduziert Kurzeinführung in CODAP: https://youtu.be/2z5H4anfHWM (ca. 5 min)
2-3	In diesen beiden Stunden geht es darum, dass die Schülerinnen und Schüler zu Data-Science-Experten werden, um anschließend selbständig arbeiten zu können. Zunächst kann, sofern noch nicht geschehen, Arbeitsblatt 1 besprochen werden. Die Präsentation zu Sitzung 2+3 kann als Lehrerhintergrundinformation oder zur Präsentation im Unterricht begleitend zu den Arbeitsblättern 3-6a genutzt werden. Einführung in nötige Grundbegriffe und Erwartungshaltung aufbauen Die SuS erhalten Informationen über grundlegende statistische Fachbegriffe (Infoblatt_Glossar). Die kennengelernten statistischen Grundbegriffe übertragen die SuS anhand von Arbeitsblatt 2 auf den vorliegenden Datensatz. CODAP und Analysemethoden Es werden Auswertungsmöglichkeiten bezüglich verschiedener Prozente und damit verbundener Aussagen erarbeitet. Zunächst sollten interessierende Merkmale umcodiert werden, so dass die Ausprägungen von sieben (täglich, mehrmals pro Woche, ... nie) auf zwei (z. B. häufig, selten) reduziert werden. Mehrere Unterrichtserprobungen haben gezeigt, dass dies ein sinnvolles Vorgehen ist und gleichzeitig das Modellieren thematisiert wird. Dies wird in Anleitung_CODAP_umcodieren (diese Anleitung ist im YouTube-Video umgesetzt) erklärt. Eine Differenzierungsmöglichkeit besteht hier erneut darin, motivierte SuS Auswertungen mit	Infoblatt_Glossar Arbeitsblatt_2_Begriffe PowerPoint-Präsentation Sitzung 2+3 (optional) Anleitung_CODAP_1dim Anleitung_CODAP_umcodieren (https://youtu.be/qcK_ZZsWfbQ) Anleitung_CODAP_2dim_Prozente Anleitung_CODAP_2dim_ausblenden I Arbeitsblatt_3_Verteilungen Arbeitsblatt_4_Verteilungen_II Arbeitsblatt_5_Verteilungen_III

	<p>den binären Merkmalen und zusätzlich mit den Merkmalen mit sieben Ausprägungen machen zu lassen (die Arbeitsblätter 3-5 lassen sich auf beide Arten bearbeiten).</p> <p>Je nach Kurs/Klasse können sich die SuS die Auswertungsmöglichkeiten anhand der Anleitungen selbst erarbeiten und bei den Arbeitsblättern 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. <i>Erfahrungsgemäß benötigen SuS Unterstützung beim Bearbeiten und Interpretieren der prozentualen Auswertungen mit Zeilen-, Spalten- oder Zellenprozenten. Einen Einstieg auch zur Diskussion kann Arbeitsblatt 6a bieten.</i></p> <p>Wir schlagen vor, mit der Anleitung (Anleitung_CODAP_umcodieren) in die Exploration zu starten und dann die Anleitung_CODAP_2dim-Prozente weiter zu nutzen.</p> <p>Technik</p> <p>Hier 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 Schülerbearbeitungen zu dokumentieren, bzw. zu kontrollieren.</p>	<p>Arbeitsblatt_6_Prozente</p> <p>YouTube-Video zum Kopieren des Arbeitsstands per Link: https://youtu.be/otLuX8hhtq8</p>
4	<p>Erwartungshaltung aufbauen und passende Fragen stellen</p> <p>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_6_Gruppenbildung). Jede Gruppe sollte vier SuS umfassen, damit die spätere Think-pair-share Phase gut funktioniert.</p> <p>Anhand von AB 6 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.</p>	<p>Arbeitsblatt_7_Gruppenbildung Plakate</p> <p>PowerPoint als Hintergrundinformation für Lehrkraft</p>

	Die Erfahrung hat gezeigt, dass das Stellen von geeigneten (statistischen) Fragestellungen für SuS eine Hürde darstellt, deshalb liegt ein weiterer Fokus dieser Stunde darauf, die SuS anhand von AB 6b in Kleingruppen geeignete Fragen erarbeiten zu lassen. Hierfür wird die Methode Think-Pair-Share genutzt.	Arbeitsblatt_8_Fragestellung
5-6	Projektarbeit in Kleingruppen In diesen beiden Stunden wird das Vorgehen der Datenexploration von den Kleingruppen zunächst geplant (Arbeitsblatt 7). Anschließend findet die Datenexploration in den Kleingruppen statt (Arbeitsblatt 8) und die Präsentation der Ergebnisse wird vorbereitet. Zu Beginn der sechsten Stunde können evtl. Kriterien für gute statistische Präsentationen besprochen werden, um die Erstellung der Präsentationen vorzubereiten (leere Power-Point-Vorlage). 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.	Arbeitsblatt_7_Planung Arbeitsblatt_8_Datenanalyse Anleitung_Hinweise_Präsentation Leere PowerPoint-Vorlage für Schülerpräsentationen
7	Präsentationen der Kleingruppen In dieser Stunde finden die Präsentationen der Schülerergebnisse statt. Dabei kann immer eine Gruppe einen speziellen Feedback-Auftrag bekommen, damit eine inhaltliche Diskussion angeregt wird (Arbeitsblatt 9). Anhand von Arbeitsblatt 10, das sich gut als Hausaufgabe und Ergebnissicherung eignet, können die SuS überprüfen, ob sie eine Datenanalyse durchführen können.	PowerPoint-Präsentationen der SuS Arbeitsblatt_9_Feedback Arbeitsblatt_10_Hausaufgabe
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 (Powerpoint_Präsentation_Stunde_8). Weiter können unter anderem persönliche und gesellschaftliche Auswirkungen einer Datenexploration angesprochen werden und versucht werden, „über den Tellerrand“ zu schauen.	Powerpoint_Präsentation_Stunde_8
Ab-schluss	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	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 JIM-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/jim-pb50-unbereinigt
--------	--	--

Weitere Hinweise

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 Power-Point-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.

Zum Umgang mit CODAP:

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

Link zum Datensatz mit allen 161 Variablen:

<https://tinyurl.com/jim-pb160>

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>