"Data exploration–data detectives at work" series of lessons

Note: This document has been automatically translated using DeepL.com. As a result, some content may not be fully accurate or correctly displayed.

## Core idea

This teaching unit is about data science content for grades 8-10 (age 14+). Using a fictitious example of an online platform that wants to tailor advertising to young people, students are motivated to search for traces and patterns in a data set as data detectives in order to advise the online platform.

**Worksheets**, **PowerPoint presentations**, **instructions**, the **YOU-PB dataset**, a **list of variables,** and an **overview** of the lessons are available. The **CODAP** software ([codap.concord.org](https://codap.concord.org/app/static/dg/de/cert/index.html)), which is freely available on the Internet, is used for data analysis.

The project works with data from over 1000 students who have provided information on many variables in the leisure and media sector (YOU-PB stands for "youth media usage in the area of Paderborn”). The data set for this series of lessons is available in different versions. On the one hand, a reduced version with 50 variables is available. On the other hand, the full version with over 160 variables can be used in lessons. It is recommended to use the data set with the reduced number of variables for better clarity. Multivariate and interesting discoveries are possible here! However, the teacher has the option of differentiating according to the ability of the class/course, or to provide an internally differentiated offer for particularly able students (see below).

The link to the data set used in the project (50 variables):

[https://tinyurl.com/you-pb-50en](https://tinyurl.com/you-pb-50en )

The series of lessons comprises approx. 8 lessons. In the first lessons, the students are introduced to the data and working with CODAP. The core of the series of lessons is independent data science project work by the students in lessons 5 and 6 with presentations in lesson 7. The project work takes place in small groups in which the students work independently as data science experts with the data set. For this purpose, they assign themselves to four groups of the YOU-PB data, analogous to the customer wishes of the online platform (this assignment takes place in lesson 4):

* Customer 1 wants to promote TikTok,
* Customer 2 wants to promote LetsPlay\_YouTube videos,
* Customer 3 would like to advertise online newspapers,
* Customer 4 would like to advertise fixed game consoles.

In the last hour, there is a reflection on the procedure for data exploration and personal and social aspects can be discussed. In addition, the topic of data cleansing, which takes up a lot of time in the work of real data scientists, can be addressed here as an excursus and worked on in CODAP.

### Target group

Mathematics from grade 8 (all school types)

Computer science from grade 8 (all school types)

### Previous knowledge

Percentages

### Time scope

8 to 10 lessons of 45 minutes each

### Goals

Statistical learning objectives

* Learners explore multivariate data
* Learners carry out group comparisons
* Learners use row, column, and cell percentages and interpret the results

Computer science learning objectives

* Learners interpret data as information
* Learners gain an understanding of the structure and functioning of IT systems
* Learners use database software
* Learners discuss social responsibility and the impact of the use of data

Media goals

* Learners use digital tools to analyze data
* Learners create presentations and practice presenting

Overview of the lesson series

|  |  |  |  |
| --- | --- | --- | --- |
| **Hour** | **Topic** | **Goals** | **Material** |
| 1 | **Introduction**  This lesson introduces the project "Data detectives at work". For this purpose, the complete framework of the lesson series is presented by the teacher via **presentation 1.** In two work phases, the students are introduced to the available survey data and the data analysis in CODAP.  The document **Lesson\_1\_Overview** provides design tips for this lesson.  Students explore the data set independently using **worksheet 1.**  Didactic notes  The introduction to CODAP can be done by means of a video or demonstration by the teacher (**see lesson 1 overview**)  The document **Teacher\_Notes** contains an overview of how to analyze categorical variables with CODAP and possible student difficulties.  Supplementary option  At APPCamps, there is learning material on leaving data, which can be used as an introduction to discussing "data traces": <https://drive.google.com/file/d/1eb0_qlnKP-63H8El0OIsDUg3xXgDtVUA/view> | Students independently explore the YOU-PB data with CODAP | Lesson\_1\_Overview  Presentation\_1  Worksheet\_1  Worksheet\_1\_solution  ListOfVariables\_YOU-PBen  Brief introduction to CODAP:  Instruction\_CODAP  Teacher\_Notes |
| 2-3 | The aim of these two lessons is for students to gain initial experience as data science experts so that they can then work independently. The focus here is particularly on working with categorical data.  If not already done, worksheet 1 can be discussed first.  **Introduction to necessary basic concepts and building up expectations**  Students receive information about basic statistical terms. Using **worksheet 2,** the students transfer the basic statistical terms they have learned to the data set at hand.  **CODAP and analysis methods**  Evaluation options regarding different percentages (row percentages, column percentages and cell percentages, see **teacher's instructions**) and associated statements are developed.  The first step is to "model the data" thematically. Variables of interest are recoded so that the values are reduced from e.g. seven (daily, several times a week, ... never) to two (e.g. often, rarely). This simplifies the evaluations.  Overview of the worksheets  There are **instructions** that students can use to work out the content themselves. Alternatively, the contents of the instructions should be worked out together with the teacher. There are also **worksheets** that should be worked on and discussed in plenary with a view to the later project work. Different distributions are compared with each other in the worksheets. The worksheets are structured in stages. It is important that the correlations found only apply to the people in the sample and can therefore be easily generalized because the sample (the data) was not collected in a representative manner.  The instructions:   * Instruction 1: Students learn to recode variables from seven expressions to two expressions * Instruction 2: Students learn the different evaluation options: Row, column and cell percentages * Instruction 3: Students learn to hide data in order to carry out analyses for subgroups * Glossary info sheet: Important technical terms are briefly explained here   The worksheets:   * Worksheet 3: two binary distributions (distributions with two values) are compared with each other * Worksheet 4: a binary distribution is set in relation to a distribution with 7 values. Either recoding is used here (recommended) or precise statements must be made * Worksheet 5 (advanced): two seven-level distributions are compared with each other. The following should be recoded * Worksheet 6 (optional, e.g. bonus):   Didactic notes  **Presentation 2** for session 2+3 can be used as background information for the teacher or for presentation in class to accompany worksheets 3-6a.  Several teaching tests have shown that recoding variables (seven different variables are reduced to two) is a useful procedure and at the same time addresses modeling. This is explained in **instruction 1**.  We suggest starting the exploration with the instructions (Instructions\_CODAP\_recode) and then continuing with Instructions\_CODAP\_2dimPercentages.  One differentiation option here is again to have motivated students carry out evaluations with the binary variables and additionally with the variables with seven values (worksheets 3-5 can be completed in both ways).  Depending on the course/class, the students can work out the evaluation options themselves using the instructions and apply them to worksheets 3-5. Alternatively, the teacher can use the PowerPoint to introduce the analysis methods and the students then work on the worksheets. This is the necessary prerequisite for independent student exploration in the following lessons. Experience has shown that students need support when working on and interpreting the percentage evaluations with row, column or cell percentages. Worksheet 6 can also provide an introduction to the discussion.  **Technology**  In this lesson, students should also be shown how they can copy graphics from CODAP into a Word file or a PowerPoint presentation. Sharing the CODAP document via a link may also be a good way of documenting and checking students' work. | Students can apply statistical terminology to the YOU-PB data  Students manipulate data by re-coding the variables  Students establish relationships between two categorical variables and distinguish between row, column and cell percentages | Presentation 2  Worksheet 2  Glossary  Worksheet\_2  PowerPoint presentation 2 (optional)  Instruction 1  Instruction 2  Instruction 3  Worksheet 3  Worksheet 4  Worksheet 5  Worksheet 6 (Bonus) |
| 4 | **Build up expectations and ask appropriate questions**  In this lesson, the students are divided into small groups for the rest of the lesson and assigned to four different content areas of the data/customers of the online platform (**Worksheet\_7**). Each group should consist of four students so that the subsequent think-pair-share phase works well.  Experience has shown that posing suitable (statistical) questions for evaluating the data is a hurdle for students, which is why another focus of this lesson is to have the students work out suitable questions in small groups using **worksheet 8.** The Think-Pair-Share method is used for this. Worksheet 8 is used for a theoretical discussion of the data and the list of variables in order to build up expectations for the coming lessons. These expectations are central to the students' own data analysis and should be recorded on posters. The posters created for this purpose should be visible to all students in the classroom during the rest of the lesson.  Didactic notes  **Presentation 3** summarizes the content of the last lesson and this one.  **Teacher notes 4** provides instructions for this lesson. | Students ask suitable statistical questions | Worksheet 7  Worksheet 8  Presentation 3  Teacher notes 4  Posters |
| 5-6 | **Project work in small groups**  In these two lessons, the data exploration procedure is first planned by the small groups (**worksheet 9**). The data exploration then takes place in the small groups (**worksheet 10**) and the presentation of the results is prepared (with **presentation\_template**). At the beginning of the sixth lesson, criteria for good statistical presentations can possibly be discussed (**Anleitung\_Hinweise\_Präsentation**) in order to prepare the preparation of the presentations. In lessons 5-6, the teacher is mainly available to answer questions and make suggestions, while the students work as independently as possible with CODAP and PowerPoint.  Didactic notes  Various trials have shown that two additional double lessons and, if necessary, homework can also be used for this project phase. | Students explore the data according to their own questions  Students document their explorations in PowerPoint  Students prepare a presentation | Worksheet\_9  Worksheet\_10  Instruction\_Presentation  Template\_presentation |
| 7 | **Presentations by the small groups**  In this lesson, the presentations of the student groups take place. One group can always be given a special feedback task to stimulate a discussion of the content (**worksheet 11**). Using **worksheet 12**, which is well suited as homework and for securing results, the students can check whether they can carry out a data analysis. | Students present the results of their project work  Students give feedback on other presentations | PowerPoint presentations by the students  Worksheet\_11  Worksheet\_12 |
| 8 | **Reflect**  In this lesson, a joint reflection of the entire project takes place. The individual steps of the data analysis carried out are assigned to the stages of the PPDAC cycle (**presentation 8**)  Furthermore, the personal and social effects of data exploration can be addressed, and an attempt can be made to think "outside the box." | Students reflect on the data analysis process  Students reflect on the social impact of data exploration | Presentation 8 |
| Conclusion | **For research purposes,** we ask that students complete an anonymous survey at the end of the lesson series and give feedback on how they liked the lesson series. The link to the survey is: <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> |
| Digression  9 | **Data cleansing as detective work**  The area of data cleansing can be discussed as an excursus in a separate lesson. The CODAP environment shown opposite contains the uncleaned YOU-PB dataset. There is also a text field with explanations and initial steps on how data cleansing can be carried out with the help of CODAP. |  |  |

# Further information

Possibility of differentiation:

The series of lessons is designed to work with a "small" data set. This contains 50 variables and offers plenty of opportunities for exploration.

For particularly motivated students, it is also possible to work with the large data set instead, which contains all 160 variables that were collected in the survey. However, this requires a high level of commitment and good work with the list of variables on the part of the students!

Another possibility for differentiation is to have particularly motivated students work with the "normal" variables with all seven values. The standard case should be working with binary variables, i.e. variables that have previously been recoded by the students, as described in Instruction\_recoding in lesson 2+3.

How to deal with the tasks:

In the series of lessons, a lot of work is done with worksheets. In order to document the learning process and at the same time keep motivation high, the tasks can also be worked on directly in a PowerPoint presentation. New tasks can be worked on new slides and at the end, results can be taken from the various lessons to create the final presentation.

Participation in the survey:

Anyone who would like to take part in the survey with their class is welcome to do so. The data will be collected completely anonymously (clarify with the school management if necessary). The data will then be included in a new edition of the data set every year, which will remain accessible via a CODAP link.

Students can take part in the survey via this link (note: 161 questions!, allow sufficient time): <http://go.upb.de/JIM-Umfrage>

# Tools + Data

CODAP: Common Online Data Analysis Platform

CODAP (codap.concord.org) is a free, browser-based software designed for use from grade 3 upwards. It supports students in learning and using statistics and data science. No registration is required to use the platform.

There is a lot of learning material on working with CODAP in the toolkit on this website.

CODAP is largely available in German. If the language is not set to German, the language can be changed at the top right.

The YOU-PB data

The YOU-PB data used here comes from a Paderborn replication study of the official JIM study, which tracks these results locally and makes them available for analysis. The JIM study has been examining the media consumption of 12 to 19-year-olds every year since 1998 in order to map trends and developments in the digital behavior of young people and to derive new impulses for education and culture.

The data is available in CODAP in four variants:

We recommend variant 1 for teaching:

50 variables, up to 7 values: <https://tinyurl.com/you-pb-50en> (Long link: <https://codap.concord.org/app/static/dg/en/cert/#shared=https%3A%2F%2Fcfm-shared.concord.org%2FzU4zLbVRo8NFLd26MxI5%2Ffile.json>))

160 variables, up to 7 values: <https://tinyurl.com/you-pb-160en> (Long link: <https://codap.concord.org/app/static/dg/en/cert/#shared=https%3A%2F%2Fcfm-shared.concord.org%2FYejBWhDQN7xJsjlcciGM%2Ffile.json>)

List of variables for the 160 data set: <https://www.prodabi.de/wp-content/uploads/List_of_variables_YOU-PB-en.pdf>