

Assignment 1 - Exploratory Data Analysis

The aim of this assignment is to use *Exploratory Data Analysis* (EDA) techniques to explore the relationship among the real-world dataset “*adult*”. This dataset is available in the R package **liver**. You could find more information about this dataset here: <https://rdr.io/cran/liver/man/adult.html>. For this assignment use only the EDA techniques that we have learned so far; see the slides of week 2 or Chapter 3 of the book *Discovering knowledge in data*.

Your task is to answer the following questions and creating a report as an **R-markdown** (.Rmd file including R code). Please upload your **R-markdown** file with the **HTML** file on Canvas at the latest on **Tuesday 28 Sep 2021 at 23:59** (Amsterdam time). The total number of points assigned is 100.

1- Loading/Attaching R packages

Load/attach the *liver* and *ggplot2* packages. If it is needed, install the packages. (5 points)

2- Importing and understading the adult dataset

Import the “*adult*” dataset which is available in the *liver* package. Report a summary of the dataset by using appropriate R functions. What is the number of variables? Which type of variables are they? Are there any missing values? What would be your strategy to deal with the missing data? (15 points)

3- Using EDA to analysis the dataset

Use the EDA techniques from week 2. For more information see slides and the documents of the week 2 at Canvas or Chapter 3 of the book *Discovering knowledge in data*. Indicate which variables have an association with the target variable **income** and which variables have no obvious association with the target variable. Explain why? (50 points)

4- Writing a summary

Summarize your EDA results from the previous question, just as if you were writing a report. (15 points)

5- Creating a report

Create a report as an R-markdown which should include the R code and the results of the code and your interpretation. (15 points)

6- Bonus question: for those who are interested (30 points)

In this part, we want to use Exploratory Data Analysis to explore the *churnTel* dataset that is available in the **R** package **liver**. You could find more information about the *churnTel* dataset here: <https://rdr.io/cran/liver/man/churnTel.html>.