

# Data Cleaning

## Homework

*Examine the '[Absenteeism-data.csv](#)' carefully. Then, use the following as a guide to how you should prepare the data for further analysis:*

- Drop the 'ID' column
- Split the reasons for absence into multiple dummy variables, and then group them in the following way:
  - Group 1: Columns 1 to 14
  - Group 2: Columns 15, 16, and 17
  - Group 3: Columns 18, 19, 20, and 21
  - Group 4: Columns 22 to 28
- After you've done that, don't forget to drop the 'Reason for Absence' column.
- Extract the month value and the day of the week from the 'Date' column. Then, drop the 'Date' column as well.
- Turn the data from the 'Education' column into binary data, by mapping the value of 0 to the value of 1, and the value of 1 to the rest of the values found in this column.

*Don't forget to create checkpoints as you go. If you have worked correctly, the final version of your DataFrame should contain the same data as the one stored in the '[df-cleaned.csv](#)' file.*

Good luck!