**GCD Assignments week 2**

*Activity 1*

Fill out the survey on: <https://docs.google.com/forms/d/1nyOZsOKbDIX559mmQgLw5i68rImEzZLxUe6tXgchh_0/viewform?usp=send_form>

After filling out the survey by all students, the results can be found on:

<https://docs.google.com/spreadsheets/d/1E6PlCvPHTN3Ngit0J46pcPAF1kKnVDB66F0xCThLb0w>

*Activity 2*

Inspect the data and find Quality measure violations

*Activity 3*

Give at least 5 scripts to solve violations, including a short statement why this is the best solution. It suffices to describe the scripts in words, like ‘delete all lines where attribute x contains non-text values’.

Fix input field to only allow the types you are concerned about.

Apply a regex to only allow strings matching male (man|m|male) or female (f|v|vrouw|female). Others will be considered as not filled.

On a new survey the first option would be the best. But on an already existing data set the second one could fix it.

*Activity 4*

Which Missing Data Mechanisms (MCAR, MAR, NMAR) do you encounter? Explain.

MCAR because people are filling in random values just for the sake of it, it does not have anything to do with the survey and the voided data is usually gibberisch