**Open refine**

1. Create a \*tidy\* spreadsheet/table listing the names of Danish monarchs with their birth- and death-date and start and end of their reign. They should be sortable by year of birth. Suitable source website is for example[here](https://kongehuset.dk/monarkiet-i-danmark/kongerakken), but you can also use another source, provided you reference it. (Collaboration is welcome. Remember to attach this spreadsheet to Brightspace submission)

In my Spreadsheet, I have chosen that Gorm den gamle is the first monarch. This choice is made on the assumption, that he was born between 908 and 918. The Birth of some of the monarchs is unknown, and therefore have I chosen too but them in the order after their reign period. So, I have chosen to put them in the order of the death, for those where I do not know the exact date or year of their birth.

In the document there is written NA which means missing information.



1. Does OpenRefine alter the raw data during sorting and filtering?

Openrefine gives the opportunity to chance and modify the messy data. When chancing the data, to become tidier, openrefine saves all the changes. This means that it is possible to see each step in the log. Eventhoug the data chances, openrefine saves the raw data.

1. Fix the[interviews dataset](https://ndownloader.figshare.com/files/11502815) in OpenRefine enough to answer this question: "Which two months are reported as the most water-deprived/driest by the interviewed farmer households?"

To get the information separated, to make it tidier, I start by finding the cell named Month\_no\_water then I choose

Value.replace(“[“,””).replace(“]”,””).replace(“’”,””)

Et billede, der indeholder tekst, skærmbillede, nummer/tal, Font/skrifttype

Automatisk genereret beskrivelse

By doing so we can now see that the two months that is registered as the most water-deprived by the interviewed farmer household is November and October. But by looking closely, you will see that some of the months are mentioned twice. This is because there is a space in front of the month. This is seen on the screenshot below.

Et billede, der indeholder tekst, skærmbillede, software, display/skærm/fremvisning

Automatisk genereret beskrivelse

To make it possible to count all the months correctly, we need to remove the space in front of some of the months. This is done by using the command “Trim leading and trailing whitespace”.

Et billede, der indeholder tekst, skærmbillede, nummer/tal, Font/skrifttype

Automatisk genereret beskrivelse

By doing so it becomes possible to count the months, and by doing so it becomes clear that the months that is most water deprived is October and September.

Et billede, der indeholder tekst, skærmbillede, software, display/skærm/fremvisning

Automatisk genereret beskrivelse

By doing it the way I did, I have changed the columns, this can become a problem if I want to use the data in the future because now, I have created some columns there isn’t named and that makes it hard to see from where the data originally came from.

1. OPTIONAL Real-Data-Challenge: What are the 10 most frequent occupations (erhverv) among unmarried men and women in[1801 Aarhus](https://raw.githubusercontent.com/aarhusstadsarkiv/datasets/master/censuses/1801/census-1801-normalized.csv)? (hint: some expert judgement interpretation is necessary. As an inspiration, [check out this chapter Making a living outside marriage from the](https://academic.oup.com/book/1458/chapter/140849368) Swedish Gender and Work project of Maria Agren)