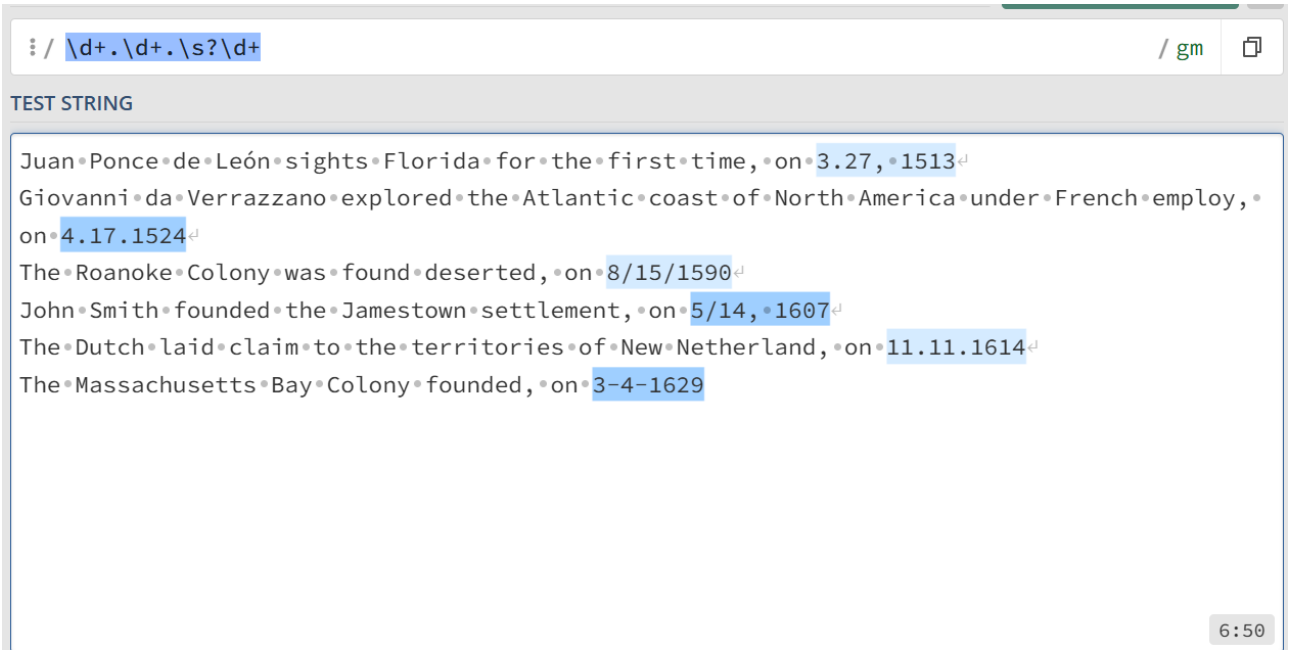
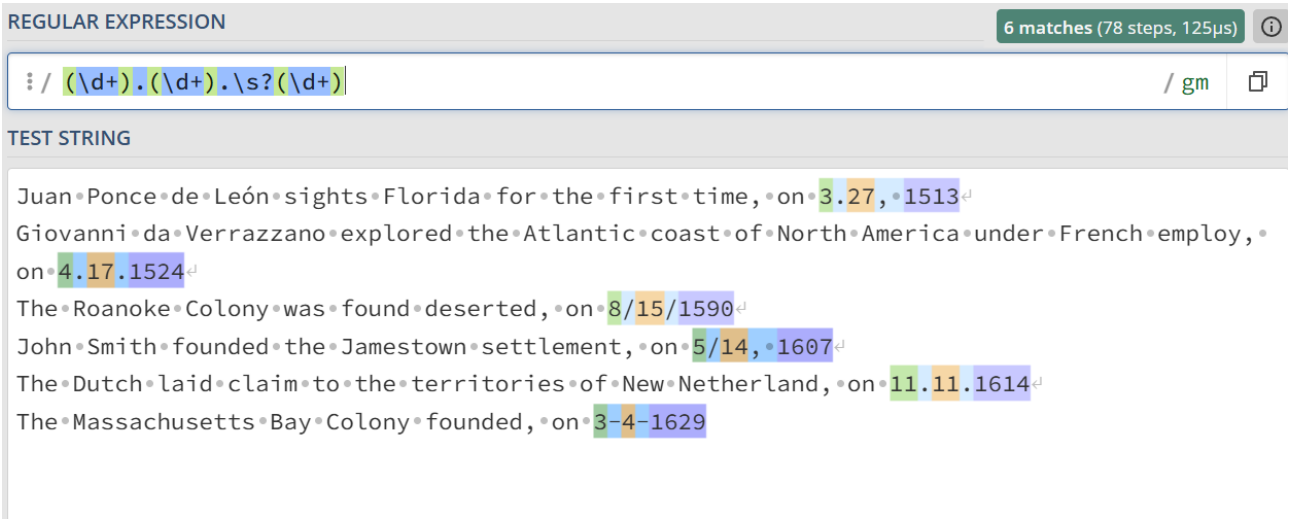
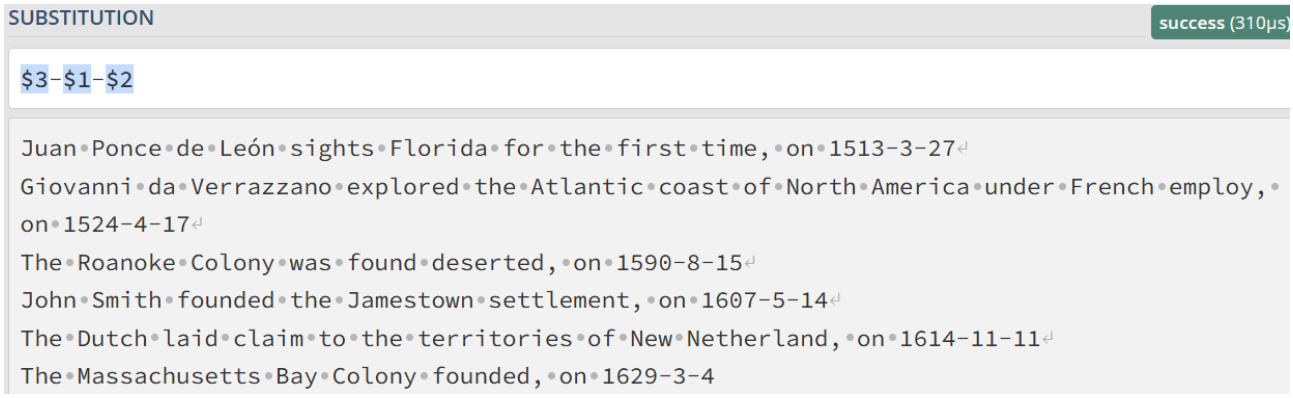
1. **What regular expressions do you use to extract all the dates in this blurb: http://bit.ly/regexexercise2 and to put them into the following format YYYY-MM-DD ?**



Ved at bruge udtrykket \d+.\d+.\s?\d+ kan finde alle datoerne

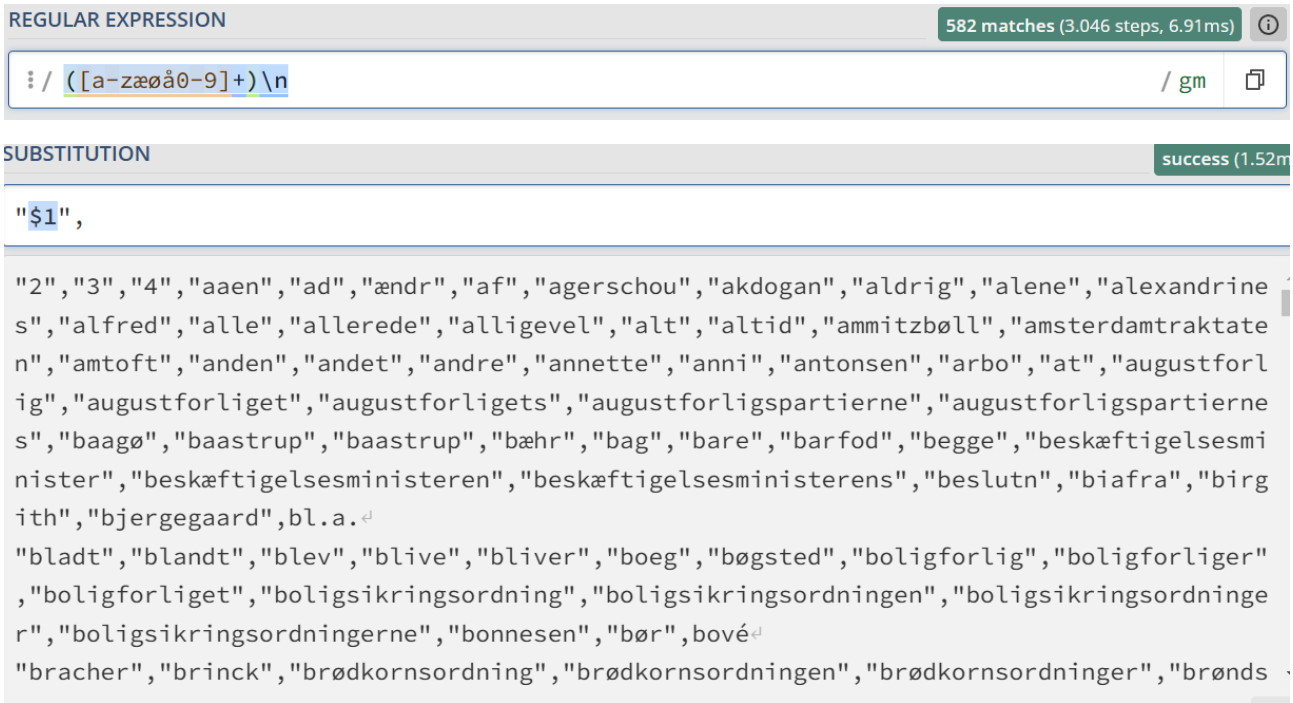


For at omstille datoerne til følgende format, YYYY-MM-DD, opdeler jeg dage måneder og år ved udtrykket: (\d+).(\d+).\s?(\d+) . Herefter skriver jeg udtrykket: $3-$1-$2 under ”substitution”. Nu vil datoerne være givet i formatet: YYYY-MM-DD

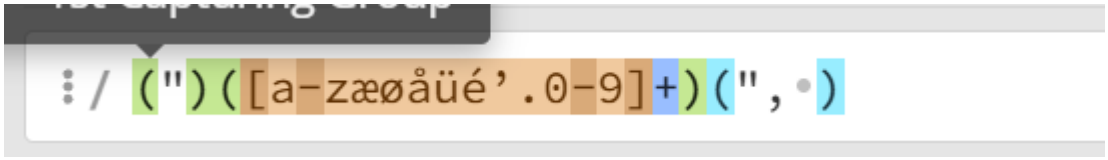


2. **Write a regular expression to convert the stopwordlist (list of most frequent Danish words) from Voyant in http://bit.ly/regexexercise3 into a neat stopword list for R (which comprises "words" separated by commas, such as http://bit.ly/regexexercise4). Then take the stopwordlist from R http://bit.ly/regexexercise4 and convert it into a Voyant list (words on separate line without interpunction)**

For at få stoplisten til at være en "Neat stopword list", skriver jeg udtrykket ([a-zæøå0-9]+)\n, herefter skriver jeg udtrykket "$1", under substitution.



For at omdanne den givet "neat stopword list" til en "Voyant list" separerer jeg tegnene fra ordene og tilføjer eventuelle bogstaver. Under Substitution skriver jeg udtrykket $2\n, som giver Voyant listen

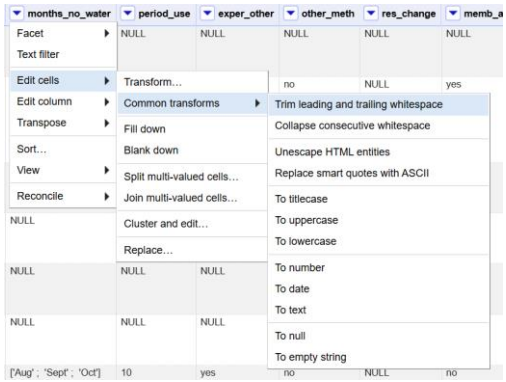


3. **Does OpenRefine alter the raw data during sorting and filtering?**

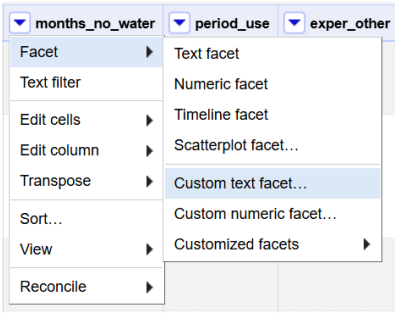
Nej, programmet vil ikke ændre det rå data under sortering og filtrering. Programmet ændrer kun hvordan dataen bliver vist.

4. **Fix the interviews dataset in OpenRefine enough to answer this question: "Which two months are reported as the most water-deprived/dryest by the interviewed farmer households?"**

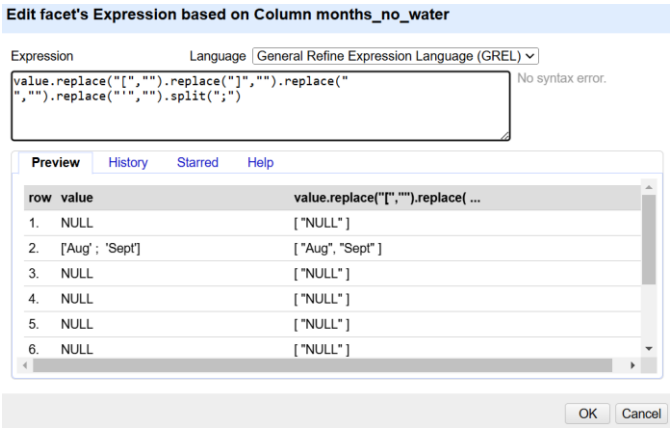
Jeg starter med at finde rækken der hedder months_no_water, hvor jeg starter med at "Trim leading and trailing whitespace"



Herefter vælger jeg Custom text facet...



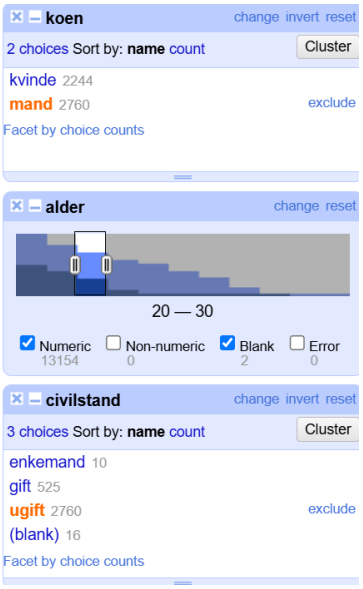
Under Custom text facet fjerner jeg de unødvnendige tegn og laver mellemrum, samt splitter ordene ad.



Herefter ændre jeg sortering af månederne efter antal og jeg kan se at oktober og september er de tørreste måneder

5. **Real-Data Challenge: What are the 10 most frequent occupations "erhverv" among unmarried men or women of 20-30 years in 1801 Aarhus census dataset? (hint: first select either men or women to shrink the dataset to a manageable size, then filter by age, and then use merging to cut the erhvervvariation ruthlessly.)**

Jeg starter med at sætte de kriterier op, som vil være for den gruppe jeg undersøger. Her er det ugifte mænd i alderen 20-30 år.



Herefter ved at sammenflette ensartede erhverv, kan jeg se at de 10 mest udbredte erhverv er:

1. Soldat
2. Tjenestekarl
3. Væver
4. Rytter
5. Matros
6. Skrædder
7. Gårdskarl
8. Land Recrut
9. Bonde
10. Læredreng