Course
CS-E4002
Course materials
Your points
Code Vault

This course has already ended.

« Preparing Data for Analysis Course materials Round 2, Exercises »

CS-E4002 / Structure and Cleaning / Filtering Data

Filtering Data

Now that the data is clean, we will see how to **filter the data**. You can, most of the time, see your data as a table with rows and columns. Cleaning it to keep only information that is relevant to you will often consist in deleting several rows and/or columns. Note that, if you haven't been able to understand the encoding of a specific row or column (e.g. you didn't have enough information on the source and can't possibly know what the comma stands for), you should get rid of it as well since it cannot be interpreted or compared to the rest of the data. With pandas, the method drop() allows you to remove parts of your dataframe to only keep what's really needed.

```
#To drop columns 1 and 4:
df.drop(['column1_name', 'column4_name'], axis=1)

#Or:
df.drop(columns=['column1_name', 'column4_name'])

#To drop lines 5 and 9:
df.drop([5, 9])
```

« Preparing Data for Analysis Course materials Round 2, Exercises »