

# clean\_up

January 8, 2024

## 0.1 Datenbereinigung des Datensatzes über das freie Rechtswissenschaftsuniversum im DACH Raum.

```
[ ]: import pandas as pd

file = "2023-12-07 Datensatz.csv"
df = pd.read_csv(file, encoding="UTF8")
```

### 0.1.1 Als erstes werden überflüssige Spalten, also solche ohne Eintrag entfernt.

```
[ ]: df.dropna(axis=1, how='all', inplace = True)
```

### 0.1.2 Dann wird manuell jeder Zeitschrift das entsprechende Land zugeordnet.

```
[ ]: unique_abbr = df['Zeitschriftkürzel'].unique()

print('\nUnique Abbreviations:')
print(unique_abbr, len(unique_abbr))
```

Unique Abbreviations:

```
['Ajbals' 'BayZR' 'GB' nan 'EJELS' 'EALR' 'FoR' 'GJLE' 'GoJIL' 'GRZ' 'HFR'
 'JIPITEC' 'KrimOJ' 'JLL' 'OdW' 'MIR' 'MIP' 'ZRI' 'Rg' 'SPWR'
 'SutdZR-WissOn' 'VLR' 'EuZ' 'ZJS' 'ZVR'] 25
```

Ajbals - Oesterreich

BayZR - Deutschland

GB - Deutschland

Digital Recht - Deutschland (Kein Kürzel)

EJELS - Oesterreich

EALR - Oesterreich

FoR - Deutschland

GJLE - Deutschland

GoJIL - Deutschland

GRZ - Deutschland  
HFR - Deutschland  
JIPITEC - Deutschland  
KrimOJ - Deutschland  
JLL - Niederlande (englische Artikel)  
OdW - Deutschland  
MIR - Deutschland  
MIP - Deutschland  
ZRI - Deutschland  
Rg - Deutschland  
SPWR - Liechtenstein  
SutdZR-WissOn - Deutschland (Achtung, wenn man Studentisches Forum für Recht und Gesellschaft googelt, findet man eine schweizer Zeitschrift. Diese ist allerdings aus Heidelberg)  
VLR - Oesterreeich  
EuZ - Schweiz  
ZJS - Deutschland  
ZIS - Deutschland (Zeitschrift für internationale Strafrechtsdogmatik - hat leider auch ZJS als Kürzel im Datensatz eingetragen)  
ZVR - Deutschland (steht nirgendow exklusiv aber klingt so)  
Zeitschrift für freiberufliche Künstler und Publizisten im Recht - Deutschland (Kein Kürzel)

### 0.1.3 Jetzt erstmal gefundenen Fehler verbessern:

1. Zeitschrift für internationale Strafrechtsdogmatik richtiges Kürzel eintragen.

```
[ ]: df[df['Zeitschrift'].str.contains('Zeitschrift für Internationale_
↪Strafrechtsdogmatik')][['Zeitschrift', 'Zeitschriftkürzel']]

[ ]:
Zeitschrift Zeitschriftkürzel
7320 Zeitschrift für Internationale Strafrechtsdogm... ZJS
7321 Zeitschrift für Internationale Strafrechtsdogm... ZJS
7322 Zeitschrift für Internationale Strafrechtsdogm... ZJS
7323 Zeitschrift für Internationale Strafrechtsdogm... ZJS
7324 Zeitschrift für Internationale Strafrechtsdogm... ZJS
...
8909 Zeitschrift für Internationale Strafrechtsdogm... ZJS
8910 Zeitschrift für Internationale Strafrechtsdogm... ZJS
8911 Zeitschrift für Internationale Strafrechtsdogm... ZJS
8912 Zeitschrift für Internationale Strafrechtsdogm... ZJS
```

8913 Zeitschrift für Internationale Strafrechtsdogm...

ZJS

[1594 rows x 2 columns]

```
[ ]: df.loc[df['Zeitschrift'].str.contains('Zeitschrift für Internationale_
↳Strafrechtsdogmatik'), 'Zeitschriftkürzel'] = 'ZIS'

df[df['Zeitschrift'].str.contains('Zeitschrift für Internationale_
↳Strafrechtsdogmatik')][['Zeitschrift', 'Zeitschriftkürzel']]
```

```
[ ]:                                     Zeitschrift Zeitschriftkürzel
7320 Zeitschrift für Internationale Strafrechtsdogm...      ZIS
7321 Zeitschrift für Internationale Strafrechtsdogm...      ZIS
7322 Zeitschrift für Internationale Strafrechtsdogm...      ZIS
7323 Zeitschrift für Internationale Strafrechtsdogm...      ZIS
7324 Zeitschrift für Internationale Strafrechtsdogm...      ZIS
...
8909 Zeitschrift für Internationale Strafrechtsdogm...      ZIS
8910 Zeitschrift für Internationale Strafrechtsdogm...      ZIS
8911 Zeitschrift für Internationale Strafrechtsdogm...      ZIS
8912 Zeitschrift für Internationale Strafrechtsdogm...      ZIS
8913 Zeitschrift für Internationale Strafrechtsdogm...      ZIS
```

[1594 rows x 2 columns]

```
[ ]: Länderkürzel = "Ajbals - Österreich,BayZR - Deutschland,GB -_
↳Deutschland,Digital - Recht - Deutschland,(Kein Kürzel)EJELS -_
↳Österreich,EALR - Österreich,FoR - Deutschland,GJLE - Deutschland,GoJIL -_
↳Deutschland,GRZ - Deutschland,HFR - Deutschland,JIPITEC - Deutschland,KrimOJ_
↳- Deutschland,JLL - Niederlande,(englische Artikel)OdW - Deutschland,MIR -_
↳Deutschland,MIP - Deutschland,ZRI - Deutschland,Rg - Deutschland,SPWR -_
↳Liechtenstein,SutdZR-WissOn - Deutschland,(Achtung, wenn man Studentisches_
↳Forum für Recht und Gesellschaft googelt, findet man eine schweizer_
↳Zeitschrift. Diese ist allerdings aus Heidelberg)VLR - Österreich,EuZ -_
↳Schweiz,ZJS - Deutschland,ZIS - Deutschland,(Zeitschrift für internationale_
↳Strafrechtsdogmatik - hat leider auch ZJS als Kürzel im Datensatz_
↳eingetragen)ZVR - Deutschland,(steht nirgendow exklusiv aber klingt_
↳so)Zeitschrift für freiberufliche Künstler und Publizisten im Recht -_
↳Deutschland (Kein Kürzel)"

# Remove parts of the string within brackets
output_string = ''
inside_brackets = False

for char in Länderkürzel:
    if char == '(':
        inside_brackets = True
```

```

elif char == ')':
    inside_brackets = False
elif not inside_brackets:
    output_string += char

output_string

output_list = output_string.split(',')

final_list = []
for entry in output_list:
    kv = entry.split(' - ')
    final_list.append(kv)

final_list[3] = ['Digital - Recht', 'Deutschland']

final_list

```

```

[ ]: [['Aj Bals', 'Österreich'],
      ['BayZr', 'Deutschland'],
      ['GB', 'Deutschland'],
      ['Digital - Recht', 'Deutschland'],
      ['EJELS', 'Österreich'],
      ['EALR', 'Österreich'],
      ['FoR', 'Deutschland'],
      ['GJLE', 'Deutschland'],
      ['GoJIL', 'Deutschland'],
      ['GRZ', 'Deutschland'],
      ['HFR', 'Deutschland'],
      ['JIPITEC', 'Deutschland'],
      ['KrimOJ', 'Deutschland'],
      ['JLL', 'Niederlande'],
      ['OdW', 'Deutschland'],
      ['MIR', 'Deutschland'],
      ['MIP', 'Deutschland'],
      ['ZRI', 'Deutschland'],
      ['Rg', 'Deutschland'],
      ['SPWR', 'Liechtenstein'],
      ['SutdZR-WissOn', 'Deutschland'],
      ['VLR', 'Österreich'],
      ['EuZ', 'Schweiz'],
      ['ZJS', 'Deutschland'],
      ['ZIS', 'Deutschland'],
      ['ZVR', 'Deutschland'],
      ['Zeitschrift für freiberufliche Künstler und Publizisten im Recht',
       'Deutschland ']]

```

Jetzt die Ländernamen den einzelnen Einträgen zuordnen

```
[ ]: import numpy as np

df['Land'] = '-'

for entry in final_list:
    df.loc[df['Zeitschrift'].str.contains(entry[0]) | df['Zeitschriftkürzel'].str.
    ↪contains(entry[0]), 'Land'] = entry[1]

df.head()
```

```
[ ]:
      Autor ohne Grad Autor mit Grad \
0      Abdulelah Alsahli      NaN
1      Endri Papajorgji, Artela Roshi      NaN
2      Jonida Gjuzi-Bushi, Ema Kristo      NaN
3      Lindita Kiri      NaN
4      Lena Hoffman, Endri Papajorgji, ,      NaN

      Titel Untertitel \
0      Challenges towards selecting Saudi Arabia as a...      NaN
1      ECJ Case 27/76 United Brands v Commission and ...      NaN
2      Didactics of intercultural communication and b...      NaN
3      Demographic transition - Global population pat...      NaN
4      About the Journal      NaN

      Zeitschrift Zeitschriftkürzel \
0      Academic Journal of Business, Administration, ...      Ajbals
1      Academic Journal of Business, Administration, ...      Ajbals
2      Academic Journal of Business, Administration, ...      Ajbals
3      Academic Journal of Business, Administration, ...      Ajbals
4      Academic Journal of Business, Administration, ...      Ajbals

      Volume Issue Datum Seitennummer \
0      Vol. 9 No. 2 - July 2023      NaN      NaN      NaN
1      Vol. 9 No. 2 - July 2023      NaN      NaN      NaN
2      Vol. 9 No. 2 - July 2023      NaN      NaN      NaN
3      Vol. 9 No. 2 - July 2023      NaN      NaN      NaN
4      Vol. 9 No. 1 - March 2023      NaN      NaN      NaN

      PDF vom Artikel (Druckversion) (Link) \
0      https://iipcccl.org/wp-content/uploads/2023/09/...
1      https://iipcccl.org/wp-content/uploads/2023/09/...
2      https://iipcccl.org/wp-content/uploads/2023/09/...
3      https://iipcccl.org/wp-content/uploads/2023/09/...
4      https://iipcccl.org/wp-content/uploads/2023/06/...

      PDF vom Artikel (Digitalversion) (Link) DOI vom Artikel \
0      NaN      NaN
```

1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

Bildschirmansicht vom Artikel (Link) \	
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN

Abstract von der Gesamtausgabe (Bildschirmansicht) (Link) \	
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN

PDF_Gesamtausgabe_Link_Druck Abstract vom Artikel Rubrik \			
0	NaN	NaN	NaN
1	NaN	NaN	NaN
2	NaN	NaN	NaN
3	NaN	NaN	NaN
4	NaN	NaN	NaN

Gesamtausgabe (Bildschirmansicht) (Link)		Land
0	NaN	Österreich
1	NaN	Österreich
2	NaN	Österreich
3	NaN	Österreich
4	NaN	Österreich

```
[ ]: df[df['Land'].str.contains('-')].count()
```

```
[ ]: Autor ohne Grad 0
      Autor mit Grad 0
      Titel 0
      Untertitel 0
      Zeitschrift 0
      Zeitschriftkürzel 0
      Volume 0
      Issue 0
      Datum 0
      Seitennummer 0
      PDF vom Artikel (Druckversion) (Link) 0
      PDF vom Artikel (Digitalversion) (Link) 0
```

DOI vom Artikel	0
Bildschirmansicht vom Artikel (Link)	0
Abstract von der Gesamtausgabe (Bildschirmansicht) (Link)	0
PDF_Gesamtausgabe_Link_Druck	0
Abstract vom Artikel	0
Rubrik	0
Gesamtausgabe (Bildschirmansicht) (Link)	0
Land	0
dtype: int64	