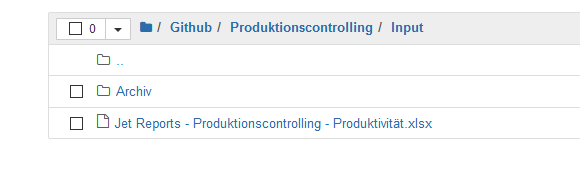
Datenprojekte

# Datenquellen (Excel, CSV, Webabruf, SQL,…)



# Import libraries

**import** pandas **as** pd #Daten

**from** matplotlib **import** pyplot **as** plt *# plots*

*#import matplotlib as mpl*

from matplotlib.ticker import FuncFormatter *# Numberformat plot ticks*

*#import numpy as np*

*# performance*

**from** datetime **import** datetime

**from** datetime **import** date *# todays date*

**import** os

now = datetime.now()

## df reduzieren – Datumsfilter

ts = pd.to\_datetime('2020-10-01', utc=True)

data = data.loc[data.Date\_reported >= ts, :]

# merge df

# how='outer','inner','left','right'

df\_I\_EW = pd.merge(df\_I, df\_EW, on='Bundesland', how='outer')

# 1. LEFT Merge

Keep every row in the left dataframe. Where there are missing values of the “on” variable in the right dataframe, add empty / NaN values in the result.

Image for post

# 2. RIGHT Merge

To perform the right merge, we just repeat the code above by simply changing the parameter of how from left to right.

# 3. INNER Merge

Pandas uses “inner” merge by default. This keeps only the common values in both the left and right dataframes for the merged data.

In our case, only the rows that contain use\_id values that are common between **user\_usage** and **user\_device** remain in the merged data — **inner\_merge**.

# 4. OUTER Merge

Finally, we have “outer” merge.

The “outer” merge combines all the rows for left and right dataframes with NaN when there are no matched values in the rows.

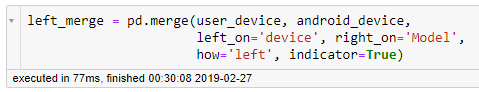
Image for post

# Merge Dataframes with Different Column Names

So we’ve talked about how to merge data using different ways — left, right, inner, and outer.

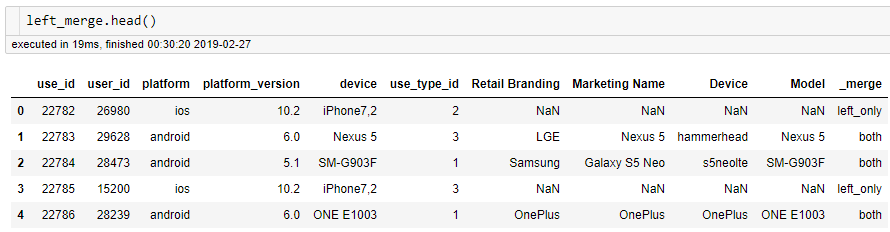
But the method on only works for the same column name in the left and right dataframes.

Therefore, we use left\_on and right\_on to replace the method on as shown below.



LEFT Merge for dataframes with different columns names

Here we’ve merged **user\_device**with **android\_device** since they both contain common codes in their columns — device and Model respectively.



LEFT Merge for dataframes with different column names