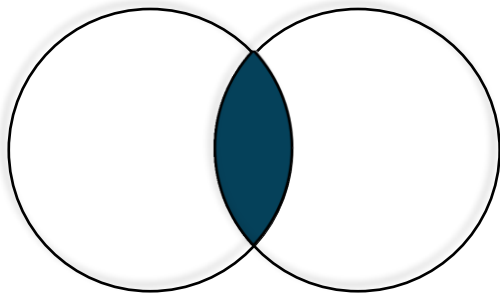


# Pandas zur Datenanalyse

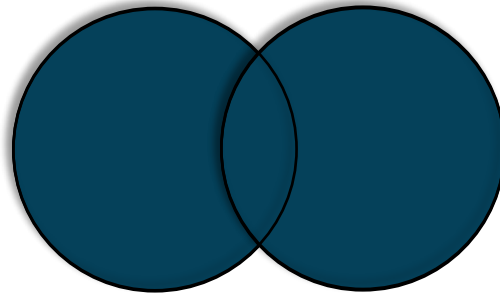
## Joins

# Verknüpfungen von DataFrames in Python

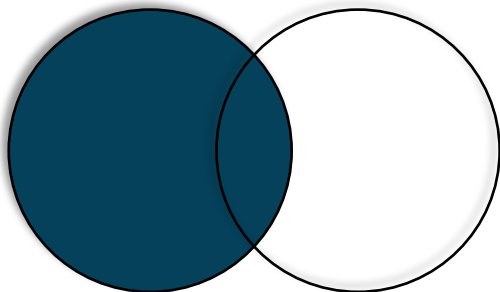
**Intern**



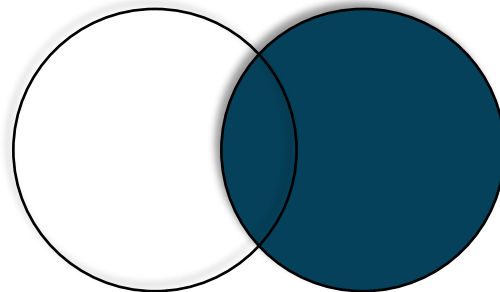
**Vollständig  
außen**



**Links**



**Rechts**



**Quelle**

Pandas PyData (2023): `pandas.DataFrame.merge`

## Beispiel

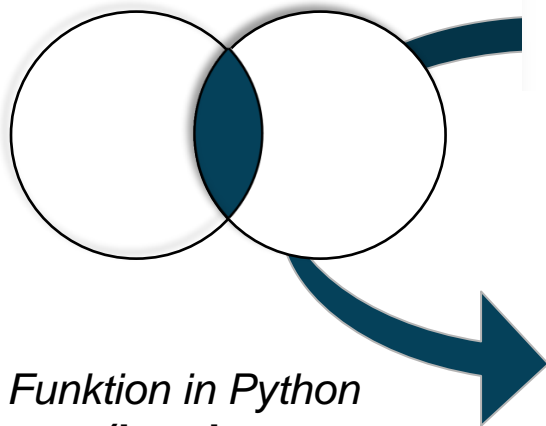
Kunden-ID	Vorname	Nachname	Stadt
1	Max	Mustermann	Berlin
2	Anna	Musterfrau	Madrid
3	Moritz	Schmidt	Paris
4	Lena	Müller	Budapest

5	Jens	Meier	Hannover
6	Mia	Hoffmann	München

Kunden-ID	Umsatz	Datum
1	123,45	01.01.2024
2	678,12	03.02.2024
3	132,54	04.05.2024
4	154,36	02.04.2024

7	432,89	07.08.2024
8	576,39	09.09.2024

## Beispiel



*Funktion in Python*  
**`pd.merge(kunden, umsatz,  
on=„Kunden-ID“, how=„inner“)`**

Kunden-ID	Vorname	Nachname	Stadt
1	Max	Mustermann	Berlin
2	Anna	Musterfrau	Madrid
3	Moritz	Schmidt	Paris
4	Lena	Müller	Budapest
5	Jens	Meier	Hannover
6	Mia	Hoffmann	München

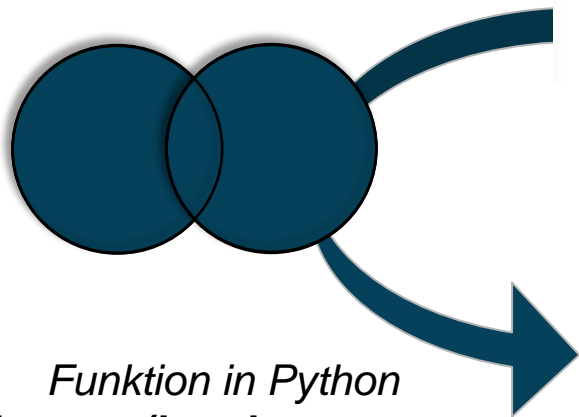
Kunden-ID	Umsatz	Datum
1	123,45	01.01.2024
2	678,12	03.02.2024
3	132,54	04.05.2024
4	154,36	02.04.2024
7	432,89	07.08.2024
8	576,39	09.09.2024

Kunden-ID	Vorname	Nachname	Stadt	Umsatz	Datum
1	Max	Mustermann	Berlin	123,45	01.01.2024
2	Anna	Musterfrau	Madrid	678,12	03.02.2024
3	Moritz	Schmidt	Paris	132,54	04.05.2024
4	Lena	Müller	Budapest	154,36	02.04.2024

### Quelle

Microsoft (2023a): Innere Verknüpfung

## Beispiel



*Funktion in Python*  
**`pd.merge(kunden, umsatz,  
on=„Kunden-ID“, how=„outer“)`**

### Quelle

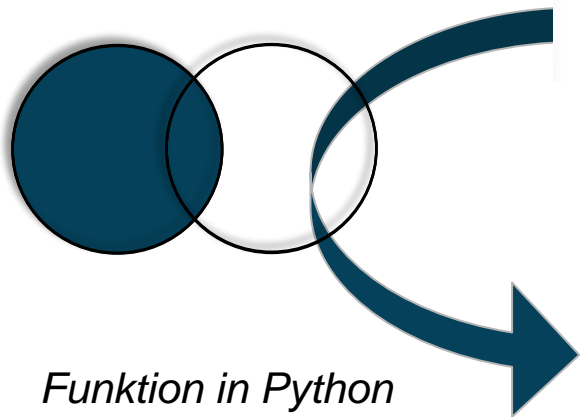
Microsoft (2023b): Vollständiger Join

Kunden-ID	Vorname	Nachname	Stadt
1	Max	Mustermann	Berlin
2	Anna	Musterfrau	Madrid
3	Moritz	Schmidt	Paris
4	Lena	Müller	Budapest
5	Jens	Meier	Hannover
6	Mia	Hoffmann	München

Kunden-ID	Umsatz	Datum
1	123,45	01.01.2024
2	678,12	03.02.2024
3	132,54	04.05.2024
4	154,36	02.04.2024
7	432,89	07.08.2024
8	576,39	09.09.2024

Kunden-ID	Vorname	Nachname	Stadt	Umsatz	Datum
1	Max	Mustermann	Berlin	123,45	01.01.2024
2	Anna	Musterfrau	Madrid	678,12	03.02.2024
3	Moritz	Schmidt	Paris	132,54	04.05.2024
4	Lena	Müller	Budapest	154,36	02.04.2024
5	Jens	Meier	Hannover	(null)	(null)
6	Mia	Hoffmann	München	(null)	(null)
7	(null)	(null)	(null)	432,89	07.08.2024
8	(null)	(null)	(null)	576,39	09.09.2024

## Beispiel



*Funktion in Python*  
**pd.merge(kunden, umsatz,  
on=„Kunden-ID“, how=„left“)**

Kunden-ID	Vorname	Nachname	Stadt
1	Max	Mustermann	Berlin
2	Anna	Musterfrau	Madrid
3	Moritz	Schmidt	Paris
4	Lena	Müller	Budapest
5	Jens	Meier	Hannover
6	Mia	Hoffmann	München

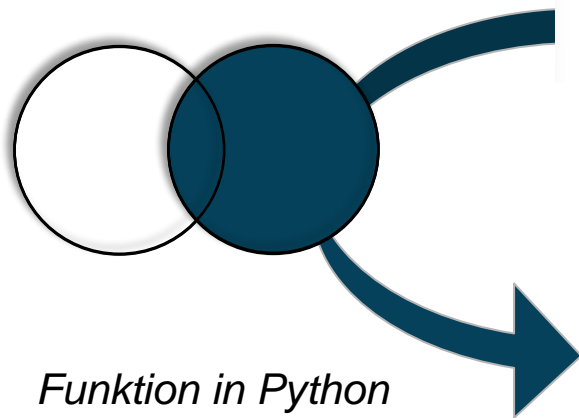
Kunden-ID	Umsatz	Datum
1	123,45	01.01.2024
2	678,12	03.02.2024
3	132,54	04.05.2024
4	154,36	02.04.2024
7	432,89	07.08.2024
8	576,39	09.09.2024

Kunden-ID	Vorname	Nachname	Stadt	Umsatz	Datum
1	Max	Mustermann	Berlin	123,45	01.01.2024
2	Anna	Musterfrau	Madrid	678,12	03.02.2024
3	Moritz	Schmidt	Paris	132,54	04.05.2024
4	Lena	Müller	Budapest	154,36	02.04.2024
5	Jens	Meier	Hannover	(null)	(null)
6	Mia	Hoffmann	München	(null)	(null)

### Quelle

Microsoft (2023c): Linke äußere Verknüpfung

## Beispiel



Kunden-ID	Vorname	Nachname	Stadt
1	Max	Mustermann	Berlin
2	Anna	Musterfrau	Madrid
3	Moritz	Schmidt	Paris
4	Lena	Müller	Budapest
5	Jens	Meier	Hannover
6	Mia	Hoffmann	München

Kunden-ID	Umsatz	Datum
1	123,45	01.01.2024
2	678,12	03.02.2024
3	132,54	04.05.2024
4	154,36	02.04.2024
7	432,89	07.08.2024
8	576,39	09.09.2024

Kunden-ID	Vorname	Nachname	Stadt	Umsatz	Datum
1	Max	Mustermann	Berlin	123,45	01.01.2024
2	Anna	Musterfrau	Madrid	678,12	03.02.2024
3	Moritz	Schmidt	Paris	132,54	04.05.2024
4	Lena	Müller	Budapest	154,36	02.04.2024
7	(null)	(null)	(null)	432,89	07.08.2024
8	(null)	(null)	(null)	576,39	09.09.2024

*Funktion in Python*

**`pd.merge(kunden, umsatz,  
on=„Kunden-ID“, how=„right“)`**

### Quelle

Microsoft (2023d): Rechte äußere Verknüpfung

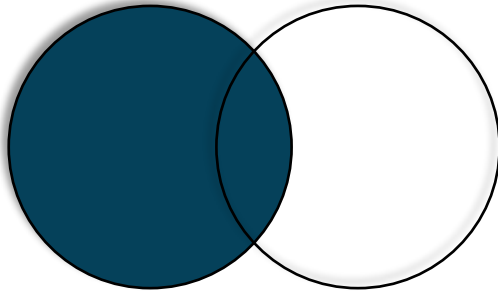
## ANTI JOIN

- Neben dem **LEFT JOIN** und **RIGHT JOIN** gibt es auch die **ANTI JOINS**
- **LEFT ANTI JOIN** analog zum **LEFT JOIN**
  - Holt sich allerdings **nicht** die **übereinstimmenden Zeilen** der **rechten Tabelle**
- **RIGHT ANTI JOIN** analog zum **RIGHT JOIN**
  - Holt sich allerdings **nicht** die **übereinstimmenden Zeilen** der **linken Tabelle**

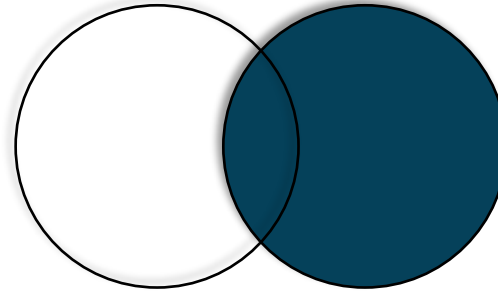


## Joinarten

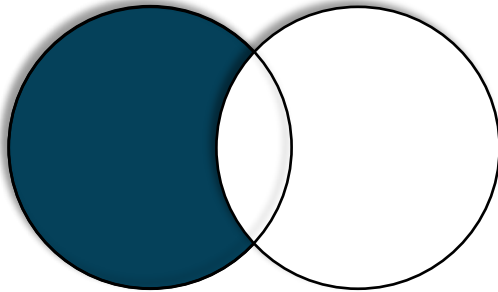
**LEFT JOIN**



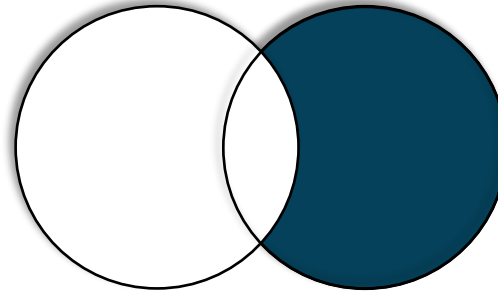
**RIGHT JOIN**



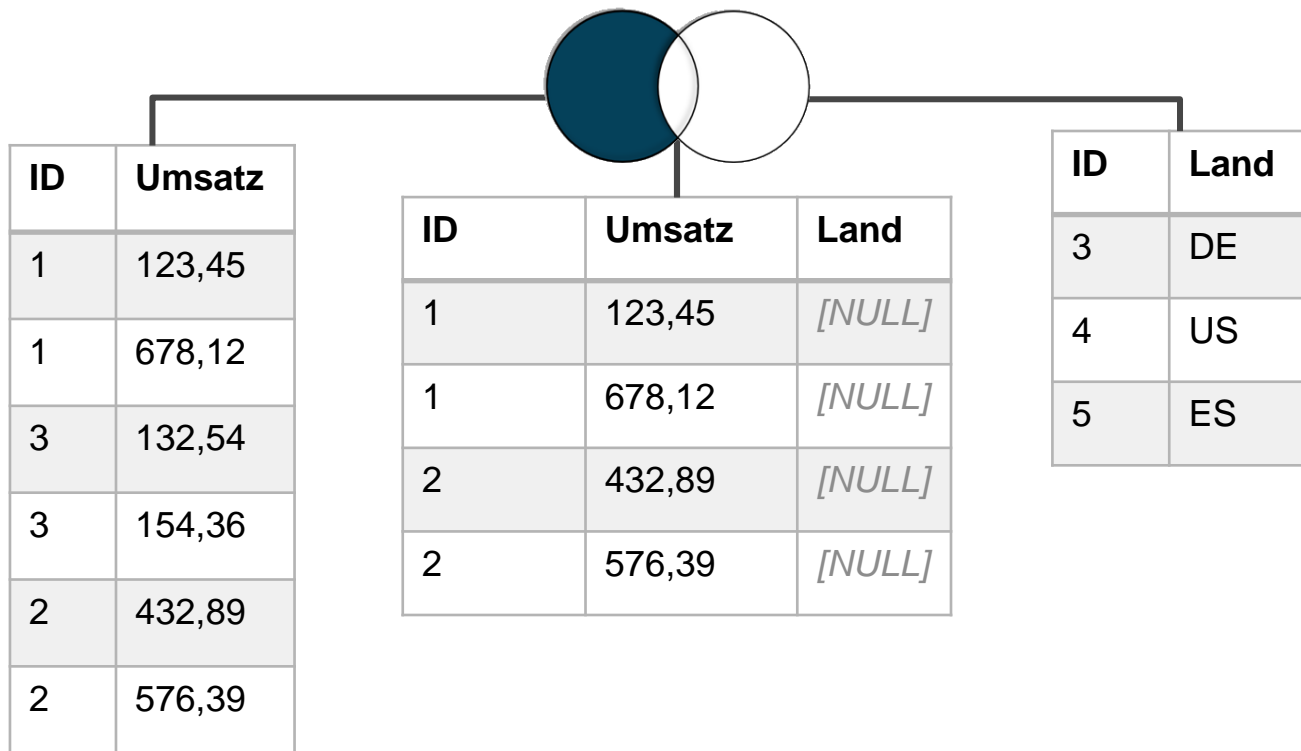
**LEFT ANTI JOIN**



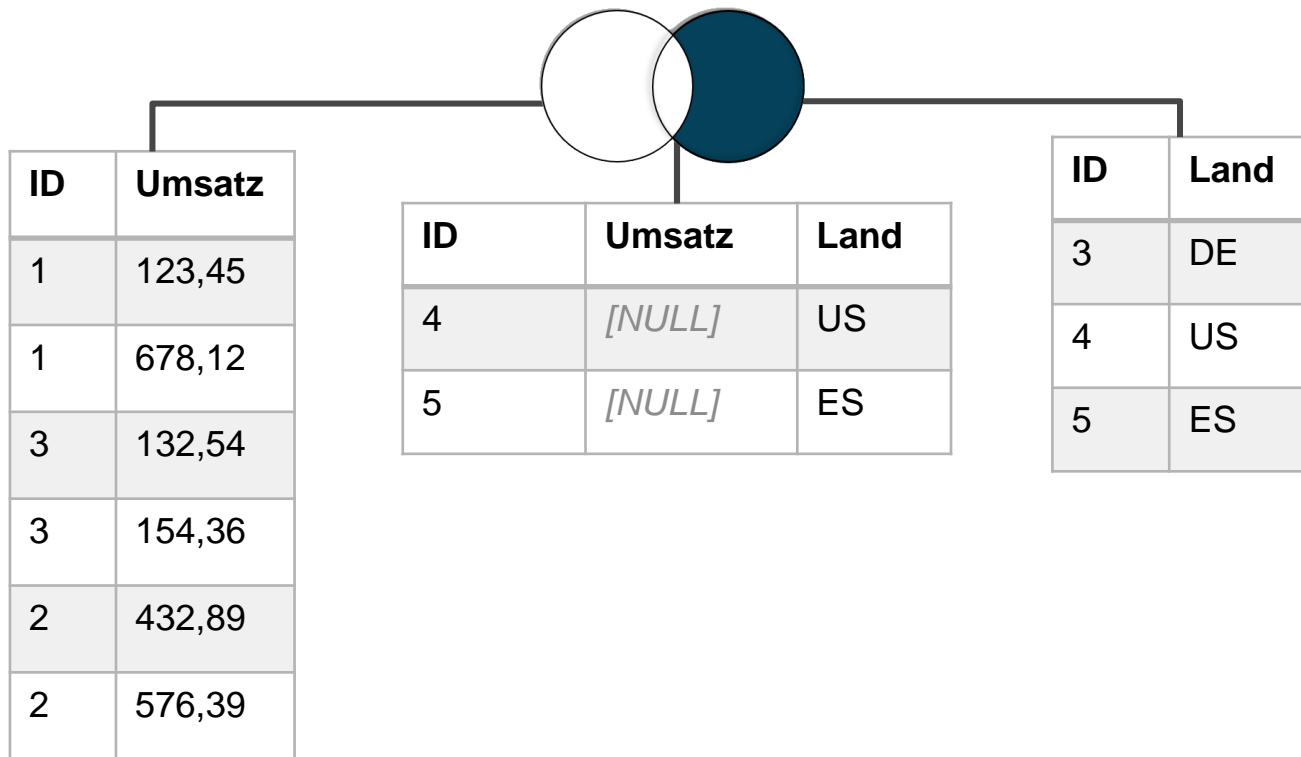
**RIGHT ANTI JOIN**



## Beispiel



## Beispiel



## Literatur

- **Microsoft (2023a):** Innere Verknüpfung, Online verfügbar unter <https://learn.microsoft.com/de-de/power-query/merge-queries-inner>
- **Microsoft (2023b):** Vollständiger äußerer Join, Online verfügbar unter <https://learn.microsoft.com/de-de/power-query/merge-queries-full-outer>
- **Microsoft (2023c):** Linke äußere Verknüpfung, Online verfügbar unter <https://learn.microsoft.com/de-de/power-query/merge-queries-left-outer>
- **Microsoft (2023d):** Rechte äußere Verknüpfung, Online verfügbar unter <https://learn.microsoft.com/de-de/power-query/merge-queries-right-outer>
- **Pandas PyData (2023):** `pandas.DataFrame.merge`, Online verfügbar unter <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.merge.html>