

WIRTSCHAFTSINFORMATIK 2

PANDAS DATAFRAMES TEIL 2

PROF. DR. BERND BLÜMEL, PROF. DR. CHRISTIAN
BOCKERMANN, PROF. DR. VOLKER KLINGSPOR

HOCHSCHULE BOCHUM

WINTERSEMESTER 2024/2025

Was Sie letzte Woche gelernt haben sollten:

- Gruppierung von DataFrames (**groupby**)
- .., dass der Umgang mit Zeit nicht ganz trivial ist

Datum und Zeit

27.11.2024 um 14:00 Uhr (Europa/Berlin) → UTC+1

Datum und Zeit

27.11.2024 um 14:00 Uhr (Europa/Berlin) → UTC+1



1732712400 Sekunden nach 1.1.1970 0:00

Datum und Zeit

27.11.2024 um 14:00 Uhr (Europa/Berlin) → UTC+1



1732712400 Sekunden nach 1.1.1970 0:00

Datumsformate / Schreibweisen

27. November 2024 14:00

2024-11-27 14:00:00

11/27/2024 2:00 pm

Datum einlesen mit Datumsformaten

Datumstext: 11/27/2024 2:00 pm

- `to_datetime` versucht Datumsformat zu raten
- Explizite Angabe des Datumsformats möglich

```
text = "11/27/2024 2:00 pm"  
dt = pd.to_datetime(text, format="%m/%d/%Y %H:%M %p")
```

→ https://pandas.pydata.org/docs/reference/api/pandas.to_datetime.html

Gruppieren und Aggregieren

```
url = "https://data.hsbo.de/kunden.csv"  
daten = pd.read_csv(url)
```

| KundeNr | Kundengruppe | Altersgruppe | Artikel | Betrag |
|---------|--------------|--------------|---------|--------|
| 282 | Schulkind | 6-10 | 4 | 0.15 |
| 269 | Mitarbeiter | 41-50 | 2 | 0.1 |
| 261 | KigaKind | 0-5 | 2 | 0.05 |
| 280 | Schulkind | 6-10 | 3 | 0.1 |

- Was ist der Gesamtumsatz je Kundengruppe **KigaKind**?

Besprechung der Wochenaufgabe



Wieviel Kuchen?

Beispiel: Sie betreiben ein kleines Café

| BonNr | Datum | Apfelkuchen | Kirschkuchen | Zitronenkuchen | Sacher |
|-------|------------|-------------|--------------|----------------|--------|
| 77 | 2023-10-12 | 0 | 0 | 0 | 0 |
| 78 | 2023-10-12 | 0 | 0 | 0 | 0 |
| 79 | 2023-10-12 | 1 | 0 | 0 | 0 |
| 80 | 2023-10-12 | 0 | 0 | 0 | 0 |

- An welchen Wochentagen müssen Sie wieviel Kuchen bestellen?
- Gibt es Tage, wo kaum Kuchen gegessen wird?

Kuchen je Wochentag

```
kuchen['Datum'] = pd.to_datetime(kuchen['Datum'])  
kuchen['Wochentag'] = kuchen['Datum'].dt.day_name()
```

| BonNr | Datum | Apfelkuchen | Kirschkuchen | Zitronenkuchen | Sacher | Wochentag |
|-------|------------|-------------|--------------|----------------|--------|------------|
| 77 | 2023-10-12 | 0 | 0 | 0 | 0 | Donnerstag |
| 78 | 2023-10-12 | 0 | 0 | 0 | 0 | Donnerstag |
| 79 | 2023-10-12 | 1 | 0 | 0 | 0 | Donnerstag |
| 80 | 2023-10-12 | 0 | 0 | 0 | 0 | Donnerstag |

→ <https://pandas.pydata.org/docs/reference/api/pandas.Series.dt.dayofweek.html>

Kuchen je Wochentag

```
nachTag = kuchen.groupby('Wochentag').sum()
```

| | Apfelkuchen | Kirschkuchen | Zitronenkuchen | Sacher |
|------------|-------------|--------------|----------------|--------|
| Dienstag | 11 | 9 | 2 | 1 |
| Donnerstag | 7 | 2 | 0 | 2 |
| Freitag | 14 | 9 | 1 | 6 |
| Mittwoch | 10 | 17 | 17 | 19 |

Kuchen je Wochentag

```
nachTag['Gesamt'] = nachTag.sum(axis=1)
```

A blue double-headed arrow labeled "axis=1" spans the width of the table. A red double-headed arrow labeled "axis=0" spans the height of the table.

| | Apfelkuchen | Kirschkuchen | Zitronenkuchen | Sacher |
|------------|-------------|--------------|----------------|--------|
| Dienstag | 11 | 9 | 2 | 1 |
| Donnerstag | 7 | 2 | 0 | 2 |
| Freitag | 14 | 9 | 1 | 6 |
| Mittwoch | 10 | 17 | 17 | 19 |

Kuchen je Wochentag

```
nachTag['Gesamt'] = nachTag.sum(axis=1)
```

| | Apfelkuchen | Kirschkuchen | Zitronenkuchen | Sacher | Gesamt |
|------------|-------------|--------------|----------------|--------|--------|
| Dienstag | 11 | 9 | 2 | 1 | 23 |
| Donnerstag | 7 | 2 | 0 | 2 | 11 |
| Freitag | 14 | 9 | 1 | 6 | 30 |
| Mittwoch | 10 | 17 | 17 | 19 | 63 |

A woman with dark hair tied back, wearing a brown jacket, sits at a dark wooden counter in a cafe. She holds a large, dark green umbrella over her head. On the counter in front of her is a small white bowl of food. The cafe has a warm, dimly lit interior with string lights and menu boards on the wall. Outside the window, it is raining heavily, and the street is blurred with lights from cars and buildings. The overall mood is quiet and contemplative.

Kuchen nach Wetterlage?

Wie beeinflusst das Wetter den Kuchenverkauf?

- Verkaufen wir mehr Kuchen bei gutem Wetter?
- Gibt es bestimmte Kuchensorten für die *gemütlich-im-Cafe-Lesenden* für schlechtes Wetter?

Wie beeinflusst das Wetter den Kuchenverkauf?

- Verkaufen wir mehr Kuchen bei gutem Wetter?
- Gibt es bestimmte Kuchensorten für die *gemütlich-im-Cafe-Lesenden* für schlechtes Wetter?
- **Woher kriegen wir Wetterdaten?**

Wie beeinflusst das Wetter den Kuchenverkauf?

- Verkaufen wir mehr Kuchen bei gutem Wetter?
- Gibt es bestimmte Kuchensorten für die *gemütlich-im-Cafe-Lesenden* für schlechtes Wetter?
- **Woher kriegen wir Wetterdaten?**

Nehmen wir einmal an, ihr Café ist in Bochum...

Open Data: **Bochumer Smart City Bänke**

- 12 Bänke in 6 Stadtteilen
- Solar-betrieben mit Handy-Ladestationen
- Viele Sensoren (**Temperatur, Regen, ...**)



Nächster Block: Datensätze verbinden (JOIN)

| Datum | Apfelkuchen | Kirschkuchen | Zitronenkuchen | Sacher |
|------------|-------------|--------------|----------------|--------|
| 2023-10-11 | 6 | 12 | 13 | 14 |
| 2023-10-12 | 4 | 0 | 0 | 2 |
| 2023-10-13 | 8 | 4 | 0 | 2 |
| 2023-10-16 | 0 | 1 | 0 | 0 |



| Datum | Regen | Temperatur |
|------------|-------|------------|
| 2024-04-12 | 0.0 | 16.07 |
| 2024-04-13 | 0.0 | 18.16 |
| 2024-04-14 | 0.0 | 14.44 |
| 2024-04-15 | 2.4 | 8.02 |
| 2024-04-16 | 5.03 | 6.65 |
| 2024-04-17 | 1.44 | 4.87 |

- Mit Pandas verschiedene DataFrames verbinden (JOIN)
- Gleiches Konzept wie in Wirtschaftsinformatik 1 (JOIN von Tabellen)

Heutige Übung

- Übungsaufgaben mit Hilfe von IBIX (Gruppieren, Filtern)
- **keine** Bonuspunkt-Aufgaben zu JOIN

Selbststudium nächste Woche

- Thema: *Daten zusammenführen (JOIN von DataFrames)*
- Foliensatz *Pandas-Join* (datascience.hs-bochum.de)
Folien voraussichtlich erst ab Donnerstag verfügbar!!
- Fragen gerne bis Dienstag per Mail an den Dozenten