

WIRTSCHAFTSINFORMATIK 2

DATENANALYSE MIT PANDAS

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

HOCHSCHULE BOCHUM

WINTERSEMESTER 2024/2025

Was Sie letzte Woche gelernt haben sollten:

- Verbinden von DataFrames (`.merge(..)`)
- .., dass das Prinzip dem von Datenbank-JOIN entspricht

Wie geht es weiter?

- Wesentliche Inhalte von Pandas abgeschlossen
- Fokus für den Rest des Semesters: Pandas nutzen um Anwendungsfälle/Aufgaben zu lösen
- jeweils Wochenaufgabe ab Donnerstag/Freitag, Lösung wird in Plenum gemeinsam entwickelt/diskutiert
- Aufgabe zu Anwendungsfall wird Teil der Klausur

Anwendungsbeispiel



A woman with long dark hair, wearing a bright yellow long-sleeved shirt, is standing at a gas station. She is holding a silver smartphone in her left hand and has her right hand on the handle of a black gas pump nozzle. The background is blurred, showing other gas pumps and a building. The text 'Günstig tanken?' is overlaid on the bottom right of the image.

Günstig tanken?

Günstig tanken – Tankpreise von *Tanker*könig

- Seit 2014 existiert Markttransparenzstelle für Kraftstoffe
- Alle Preisänderungen müssen veröffentlicht werden
- Grundlage für Tank-Apps und andere Dienste

Viele spannende Fragen:

Wo tankt man in Bochum am preiswertesten?

Welcher Wochentag ist am günstigsten?

Tankpreise

- Jede Preisänderung mit Datum, Uhrzeit und Tankstelle
- Jeweils aktueller Preis für E5, E10 und Diesel

date	station_uuid	diesel	e5	e10
2020-01-01 02:03:03+01	b7d..7c7	1.289	1.409	1.369
2020-01-01 04:36:04+01	51d..03a	1.309	1.429	1.409
2020-01-01 04:36:04+01	51d..03a	1.319	1.439	1.399
2020-01-01 04:36:04+01	51d..03a	1.319	1.439	1.419
2020-01-01 04:36:04+01	cd8..5c4	1.319	1.439	1.419
2020-01-01 04:52:03+01	836..0c6	1.308	1.428	1.408
2020-01-01 05:00:05+01	ae4..09a	1.419	1.559	1.539

https://data.hsbo.de/tankpreise_Bochum_2020.csv

(Auszug Tankpreise Bochum, 2020)

Tankpreis-Daten

- Vollständiger Datensatz ist über 60 GB groß
- Zum Ausprobieren: Einschränkung auf Tankpreise in Bochum vom 1.1. bis 31.12.2020

https://data.hsbo.de/tankpreise_Bochum_2020.csv

Daten der Tankstellen

- Datei `tankstellen.csv` enthält alle bekannten Tankstellen (Stand: Ende 2022)

`https://data.hsbo.de/tankstellen.csv`

- Jede Tankstelle hat eindeutige ID (Spalte `uuid`)
- Zusätzlich Markenname, Adresse, Koordinaten

Daten der Tankstellen

uuid	name	brand	street	house_number	post_code	city	latitude	longitude
000..001	BAGeno Raiffeisen eG	nan	Künzelsauer Strasse	7	74653	Ingelfingen	49.29682	9.66138
005..d12	familia Tankstelle	FAMILA	Pascalstrasse	9	25442	Quickborn	53.74215	9.94124
005..d16	star Tankstelle	STAR	Riehler Strasse	240	50735	Köln	50.9618	6.98007
005..d1f	star Tankstelle	STAR	BAB 10 / Seeberg Ost	nan	15345	Altlandsberg	52.55016	13.68212
005..d23	star Tankstelle	STAR	Duisburger Straße	130	47166	Duisburg	51.48979	6.78373
005..d30	star Tankstelle	STAR	Hochdahler Straße	127	40724	Hilden	51.17626	6.93881
005..d33	star Tankstelle	STAR	Coswigerstraße	5	39261	Zerbst	51.96113	12.09988

Fragestellungen:

- Welcher Wochentag ist am preiswertesten?
- Wie hat sich der Durchschnittspreis über das Jahr 2020 entwickelt?
- In welchem PLZ-Bereich kann man in Bochum am günstigsten tanken?
- Wie häufig haben die Marken *Shell* und *ESSO* jeweils die Preise geändert?

Wie gehen Sie vor?

Heutige Übung

- Umgang mit IBIX + Jupyter-Notebooks
- Übungsaufgaben mit Hilfe von IBIX

Selbststudium nächste Woche

- Wochenaufgabe mit Anwendungsfall + Fragestellung
voraussichtlich erst ab Donnerstag/Freitag verfügbar!!
- Fragen gerne bis Dienstag per Mail an den Dozenten