

# Data Science

## Übungsblatt 9

### Aufgabe 1 (Einfache Datenanalysen)

Wir betrachten im Folgenden einen Lieferdienst für Lebensmittel. Der Lieferdienst führt natürlich über die bestellten Waren pro Kunde genau Buch. Kunden bestellen Waren für jeweils eine Woche im voraus, z.B. 3 Portionen Äpfel für Dienstag, Donnerstag und Sonntag, 1 Portion am Samstag.

Unter der URL

<https://data.hsbo.de/lieferdienst-bestellungen.csv>

befindet sich ein Datensatz mit Bestellungen. Die folgende Tabelle zeigt je Kunde die bestellten Artikel pro Tag. Der Wert ist jeweils als *Anzahl Portionen* gedacht.

Kunde	Woche	Kategorie	Produkt	Mo	Di	Mi	Do	Fr	Sa	So
K7	5/2021	Obst	Äpfel	3	0	0	3	0	1	3
K7	5/2021	Fertiggerichte	Nudeln	1	2	1	3	3	3	2
K4	5/2021	Fertiggerichte	Bratkartoffeln	0	3	1	2	2	1	3
K4	5/2021	Gemüse	Paprika	0	0	0	3	1	0	3
K4	5/2021	Getränke	Bier	3	2	3	1	1	1	2

1. Berechnen Sie eine Spalte **Gesamt**, die für jede Zeile die Gesamtanzahl an Portionen für ein Produkt pro Kunde enthält.
2. Was ist die größte Anzahl an Portionen die ein Kunde in einer Woche bestellt hat?
3. Die Kategorien **Gemüse** und **Obst** enthalten die kritischen Frischeprodukte, für die auf jeden Fall genügend Lagerkapazität im lokalen Warenverteilzentrum vorhanden sein muss.

Berechnen Sie die Portionenanzahlen für diese Frischeprodukte pro Kalenderwoche. Dafür bietet sich z.B. die **groupby** Funktion von **DataFrame** an.

Welche Kalenderwoche hat die meisten/wenigsten Frischeportionen? Wie sehr schwankt dieser Wert? (Standardabweichung?)

4. Berechnen Sie den Anteil der Frischeportionen an der Gesamtzahl bestellter Portionen über alle Wochen.

**Hinweis:** Für das Filtern nach Frischeprodukten ist eine ODER Verknüpfung von Bedingungen erforderlich. Dazu können zwei Bedingungen mit einem **|** verbunden werden (senkrechter Strich), was dem ODER-Operator von Pandas entspricht.