

Wirtschaftsinformatik 2

Wochenaufgabe - Woche 8

Mit `groupby(...)` steht in dieser Woche eine mächtige Möglichkeit von Pandas bereit, mit der viele spannende Fragestellungen in Daten beantwortet werden können. Im Prinzip ist das eine automatisierte Erweiterung von *Filtern und Aggregieren* aus der letzten Woche.

Aufgabe 1 (Statistiken mit Pandas DataFrame)

In dieser Aufgabe betrachten wir wieder die Daten von Fahrradkäufern aus der letzten Woche. Die Daten sind wieder unter der URL

<https://data.hsbo.de/bike-sales.csv>

verfügbar. Den Datensatz hatten Sie in der letzten Woche ja schon eingeladen und hinsichtlich einiger Aspekte untersucht. Hier ist nochmal ein Ausschnitt der Daten:

ID	Alter	Geschlecht	Einkommen	Verheiratet	Kinder	Bildung	Beruf	Pendeldistanz	AnzahlAutos	Hausbesitzer	Region	Fahrradkauf
12496	42	weiblich	40000	Ja	1	Bachelor	Handverk	0-1 Kilometer	0	Ja	Europa	Nein
24107	43	männlich	30000	Ja	3	Mittlere Reife	Angestellter	0-1 Kilometer	1	Ja	Europa	Nein
14177	60	männlich	80000	Ja	5	Mittlere Reife	Freiberufler	2-5 Kilometer	2	Nein	Europa	Nein
24381	41	männlich	70000	Nein	0	Bachelor	Freiberufler	5-10 Kilometer	1	Ja	Asien-Pazifik	Ja
25597	36	männlich	30000	Nein	0	Bachelor	Angestellter	0-1 Kilometer	0	Nein	Europa	Ja
13507	50	weiblich	10000	Ja	2	Mittlere Reife	Arbeiter	1-2 Kilometer	0	Ja	Europa	Nein

Mit diesen Daten lassen sich natürlich viele Fragestellungen untersuchen. Wir sind an den folgenden Fragen interessiert:

- Aus welcher Region stammen die meisten Kunden, die ein Fahrrad gekauft haben?
- Wie ist das durchschnittliche Einkommen und die durchschnittliche Anzahl an Kindern je Region?

Hinweis: Es ist hilfreich, sich aus den Daten zunächst einen DataFrame zu basteln, der nur die Spalten enthält, die für die Fragestellung relevant sind. Ggf. kann es hilfreich sein, zusätzliche Spalten zu berechnen oder Werte in Spalten zu ersetzen.