

1.2 Hands-On: Customer Lösung

Marco Wähler

2022-03-25

Aufgabenstellung - Lösung

Fallbeispiel: Sie haben verschiedene Daten über die Besucher*innen ihres Webshops erhoben. Erstellen Sie die Variablen, indem Sie den jeweiligen Code kopieren oder abschreiben und ausführen.

```
#Alter
age <- c(64, 23, 69, 70, 72, 27, 33, 36, 51, 74, 55, 61, 41, 41, 52)
#1 binäre Geschlechtskodierung für männlich und 2 für weiblich
sex <- c(1, 2, 1, 1, 2, 2, 1, 2, 2, 1, 1, 1, 2, 2, 1)
#0 kein kauf und 1 kauf
buy <- c(1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0)
```

Aufgabe 1

Dummerweise sind die Daten über die Besucher*innen in separaten Vektoren gespeichert. Bitte erstellen Sie einen Data Frame “customer” mit den drei Variablen age, sex und buy.

```
customer <- data.frame(age, sex, buy)
```

Aufgabe 2

Wie viele Fälle hat der neue Datensatz?

```
nrow(customer)
```

```
## [1] 15
```

Aufgabe 3

Berechnen Sie das durchschnittliche Alter der Besucher*innen

```
mean(customer$age)
```

```
## [1] 51.26667
```

Aufgabe 4

Können Sie aus dem Alter (Variable `age`) das Geburtsjahr der Besucher*innen berechnen? Zwei Hinweise:
1. Alle Daten wurden im Jahr 2020 erhoben. 2. Eine mathematische Operation übernimmt R für jedes Element in einem Vektor.

```
year_born <- 2020 - age
year_born
```

```
## [1] 1956 1997 1951 1950 1948 1993 1987 1984 1969 1946 1965 1959 1979 1979 1968
```

Aufgabe 5

Sie haben ebenfalls die Zeit (in Minuten) erhoben, die Besucher*innen auf Ihrem Webshop verbringen.

```
time <- c(24, 20, NA, 20, 13, 9, 7, 5, 11, NA, 11, 23, NA, 6, 3)
```

Die neue Variable müssen Sie ebenfalls dem Datensatz zuweisen, damit Sie den Datensatz nicht überschreiben, müssen Sie eine neue Variable hinzufügen. Überprüfen Sie zunächst, ob die Variable fehlende Werte hat. Wenn ja, wie viele?

```
sum(is.na(time))
```

```
## [1] 3
```

```
customer$time <- time
```

Aufgabe 6

berechnen Sie die durchschnittliche Zeit, die Besucher*innen auf Ihrem Webshop verbringen. Wenn das Ergebnis "NA" kommt, dann gucken Sie sich die Hilfe zur entsprechenden Funktion an.

```
# Hilfe über ?mean
mean(time, na.rm = TRUE)
```

```
## [1] 12.66667
```