

Aufgabenblatt dplyr

2022-11-11

Öffentlichkeitsarbeit bei tödlichen Polizeischüssen

1)

- Kreiere eine neue Variable “Presse”. Diese nimmt den Wert “Ja” an, falls in dem Datensatz eine Quelle angegeben ist.
- Berechne den Prozentsatz pro Bundesland, bei dem es für den Vorfall eine Quelle gibt.

2)

- Bei manchen Vorfällen ist ein Wikipedia Artikel verlinkt. Erstelle eine neue Variable, die angibt, ob es sich bei der Quelle um einen Wikipediaartikel handelt.
- Tip: Nutze die Funktion `str_detect()` aus dem package `tidyr` um die Präsenz von einem string in einer Variablen zu finden.

Berliner Wohnungsmarkt

- Lade den Datensatz *immo_reduced*

1)

- Erstelle eine neue Variable `sqm_price`, welche den Quadratmeterpreis der angebotenen Wohnungen berechnet.

2)

- Recherchiere alle Wohnungsunternehmen, welche mehr als 3.000 Wohnungen in Berlin vermieten und speichere die Namen in einem Vektor.
- Recherchiere alle kommunalen Wohnungsunternehmen in Berlin und speichere die Namen in einem Vektor.

3)

- Erstelle eine neue Variable “Vermieter_typ”, welcher die angebotenen Wohnungen nach Vermietertyp klassifiziert. Die Ausprägungen sind: Privat, von Enteignung betroffen, kommunale Unternehmen, Unternehmen. Dafür musst du herausfinden, ob ein String sich in diesem Vektor befindet.
- Nutze `str_detect` aus dem packet `stringr` um die Präsenz einer Zeichenabfolge in einem String zu finden. Unterstützung findest du hier: <https://stackoverflow.com/questions/35962426/multiple-strings-with-str-detect-r>
- Speichere den Datensatz