

# Case Study Data Engineer

## Aufgabe 1 – Scraping

Entwickle einen Scraper für [www.biermarket.de](http://www.biermarket.de) in Python gem. folgenden Aufgaben und beantworte die zugehörigen Fragen.

- a) Extrahiere alle bayrischen Biere auf "www.biermarket.de" und speichere diese als csv-Datei.
  - Nutze „requests“ und „BeautifulSoup“, um das HTML abzurufen und zu verarbeiten
  - Speichere die wichtigsten Informationen zu jedem Bier
- b) Lade die erstellte csv-Datei und verarbeite die Daten, um folgende Fragen zu beantworten.
  - Was ist das günstigste Bier nach Standard-Preis?
  - Welcher Hersteller hat durchschnittlich das beste Rating?
  - Wenn du 10 Liter Bier kaufen willst und das Pfand zurückbringst, welches Bier solltest du kaufen und wie viel musst du insgesamt ausgeben?
- c) Welche Alternativen fallen dir für die Datenerhebung im Vergleich ein? Gibt es Vor- und Nachteile anderer Methoden?

## Aufgabe 2 – Datenanalyse

Dir liegt ein Datenset mit Verkaufszahlen (eine Zeile ist ein Verkauf) vor, das verschiedene Merkmale wie Gerätetyp, Preis, Cashback, Bewertungen und weitere Attribute enthält. Finde für jede Analyse eine geeignete Visualisierung, die die relevanten Muster und Zusammenhänge klar darstellt. Nutze eine Methode, die dir geeignet erscheint, um die Daten anschaulich und verständlich zu präsentieren.

- Stelle sicher, dass deine Analysen gut dokumentiert sind, damit der Zusammenhang zwischen den Ergebnissen und den gewählten Visualisierungen nachvollziehbar ist.
  - Berücksichtige dabei auch die Qualität der Daten. Führe geeignete Vorverarbeitungs- und Bereinigungsschritte durch, bevor du mit der Analyse beginnst.
- a) **Untersuche die beliebtesten Provider und Tarife:**
    - Welche Provider haben die höchsten Verkaufszahlen?
    - Welche Tarife sind am gefragtesten?
  - b) **Untersuche die Beziehung zwischen Verkaufszahlen und dem durchschnittlichen Monatspreis, unterteilt nach Segment:**
    - Analysiere, ob und wie der durchschnittliche monatliche Preis den Umsatz beeinflusst und ob es Unterschiede je nach Transfertyp gibt.
    - Achte darauf, die Daten entsprechend zu segmentieren, um zu erkennen, ob der Preis bei unterschiedlichen Transfertypen und TV-Optionen andere Auswirkungen hat.
  - c) **Untersuche den Einfluss von Cashback auf die Verkaufszahlen:**
    - Finde heraus, ob es einen Zusammenhang zwischen dem angebotenen Cashback und den Verkaufszahlen gibt.
    - Sind höhere Cashback-Werte besser als niedrigere Monatspreise?
  - d) **Reflektion:**
    - Wie könnte man die bestehenden Analysen noch verbessern?
    - Welche Daten würdest du gerne noch anfordern, um tiefer einzusteigen?