

# Data 2 Dollar: Challenge 1

Flurin Oehler 2020 – 062 – 007

## Begründung Genutztes Verfahren

Ich habe versucht die Challenge mit Scrapy/Selenium in Visual Studio Code zu lösen. Ich habe mich jedoch im Laufe des Prozesses entschieden auf Scrapethings zu wechseln. Scrapethings habe ich aus folgenden Gründen bevorzugt:

1. Manuelle Eingabe von gewünschten Datenfelder.
2. Einfacheres navigieren auf die zweite Seite.

Ich hätte gerne Scrapy/Selenium in Visual Studio Code genutzt da ich folgende Vorteile sehe:

1. Den Code schnell zu entwickeln (durch Copilot)
2. In einem kurzen Zeitraum viele Iterationen möglich waren und der Code verfeinert werden kann
3. Ich hier eine Übertragbarkeit bzw. Anpassungsfähigkeit des Codes an ähnliche Problemstellungen sehe. Die angesprochene Übertragbarkeit und schnelle Anpassung ist das, was mir bei anderen Tools wie Scrapethings.

## Probleme mit Scrapy/Selenium

Folgende Probleme konnte ich in der mir zur Verfügung stehenden Zeit nicht lösen, weshalb ich meine Herangehensweise geändert habe:

1. Ein **Vordringen auf die zweite Seite** um auch dort Daten zu Scrapen war nicht möglich. Wobei ich hier gezielt versucht habe durch mehrere Iterationen auf die zweite Seite zu gelangen, konnte der erstellte Scraper dies nicht bewerkstelligen → in Scrapethings einfach manuell wählbar
2. In meiner Webansicht von AirBnB (nach Rückmeldung von Kommilitonen bei ihnen nicht durchgehend der Fall) erscheinen bei mir in der Mitte der 1. Seite eine Linie mit 6 Resultaten für «**in ähnlichen Zeiträumen verfügbar**». Diese wurden immer mit gescraped. Auch bei mehrfacher Iteration und durch die Definition von verschiedenen neuen Kriterien konnte ich dieses Problem nicht lösen → bei Scrapethings trat dieses Problem nicht auf
3. **Anzahl gescrapter Wohnungen**. Diese war zunächst akkurat aber durch das Hinzufügen von Kriterien und dem nicht zur Verfügung stehen der 2. Seite, wich diese dann oft ab. Dieses Problem wollte ich ggf. zum Schluss angehen sofern die Probleme 1 & 2 behoben sind.