NotSoUniversal Search Engine

From the very beginning of backpack, we wanted to build an universal product search engine, and are working hard to make that happen. In this project, you'll build a not-so-universal subset of that search engine.

Since we have to start somewhere, instead of every sites, we will scrape amazon only. And amazon has like almost quarter billion products. For the first part of the project, you are given ten thousand ASIN (Amazon Standard Identification Number) and your goal is to scrape data from their product pages in this format. We've provided some examples here for you to look at.

Think of the project in this way: you got a list of 10k most popular items for your users and you want to deploy a minimal search engine for them by tomorrow. Do whatever you need to do. Done is better than perfect! So it's better not to fret over implementing the fastest or most clever solution from the start. You can always iterate on it later.

Deliverable

By the end of the first part of the project, you should have data for (almost) all of the 10k items in the given format. Few products might be unavailable or stocked out by now, that's fine. You should also answer these questions,

- How long does it take for your program to scrape each product on avg? What could be some possible ways to make it even faster?
- Did you get flagged as robot while scraping? What % of the time? How to get around/minimize that?

You'll have 36 hours (or less if you want to finish early) to work on the first part. After that, you'll have additional 36 hours to work on the second part where you'll deploy an actual search engine!

We're Here For You

Don't waste too much time getting stuck at something. Feel free reach out to us if you have any question during the project and we'll try to get back asap. Think out

loud. Communicate early. Communicate often. Hope you have fun building this. Let the hacking begin! 🚀