Scrapers spec for NYTimes.com, WashingtonPost.com and Bloomberg.com

NYTimes.com

- Scrape all the latest articles at the top
- Ignore the world cup articles
- Ignore the Opinion section
- Ignore the Great Read
- Ignore "In case you missed it" section
- Ignore this section:

  -

- Scrape the "More News" section
- Ignore the Weather
- Ignore the Paid post
- Ignore the Well section
- Ignore the Holiday Shopping section
- Ignore the Culture and Lifestyle section
- Ignore the Sports section
- Ignore the Cooking section
- Ignore the Games section
- Scrape the hole News section
- Ignore the Opinion section
- Ignore the Arts, Living and Travel sections

WashingtonPost.com

- Scrape the top news articles
- Ignore all the following sections:
  - Opinions
  - World Cup

Listen: Black Friday is now more than a month long. What's changed and why.

Play: Today's crossword puzzles and other games

Carolyn Hax: Telling mother-in-law her scented products make me sick

Cook: No one but you will know these brothy beans started in a can

- ○
  - ○ Any paid content
  - ○ Live events
  - ○ Most read
  - ○ Holiday gift guide
  - ○ Don't miss
  - ○ Advice
  - ○ For you
  - ○ Well being
  - ○ Lifestyle
  - ○
- Scrape all the other sections not mentioned above

Bloomberg.com

1. Scrape all the articles in the top carousel
2. Scrape all the latest articles at the top
3. Ignore:
   a. Most Read
   b. Opinion
   c. Sponsored content
4. Scrape all the articles in the sections with headings like this, including Markets and Crypto:

   a.

5. Ignore:
   a. Bloomberg Quicktake
   b. Bloomberg CityLab

6. Scrape everything else

All scrapers

Scrape all the news once per hour and update the same API endpoint once per hour. Remove duplicates. At the end of each day, store the current day API endpoint containing all unique articles for that day.

With all of these scrapers I'm only interested in the news, but I would like all the news that is posted to the homepage (that's basically the only rule for now).