



Assignment for Tech Internship via Internshala

Hey there!

Congratulations! You were shortlisted from over 1800 candidates who applied for the internship.

About us

We are a startup focused on building marketing-tech products. We have a sister concern – Barsaati Films, which is our film studio arm.

About the founder

Our founder, Ankur Singla, is a tech entrepreneur who sold his Sequoia-backed company, Tapzo, to Amazon Inc in 2018. Tapzo was an all-in-one app where you could book a cab, order food, book a flight, or recharge your phone. Today, a lot of the Tapzo backend codebase lives on in Amazon systems powering multiple Amazon Pay categories.

What are we building

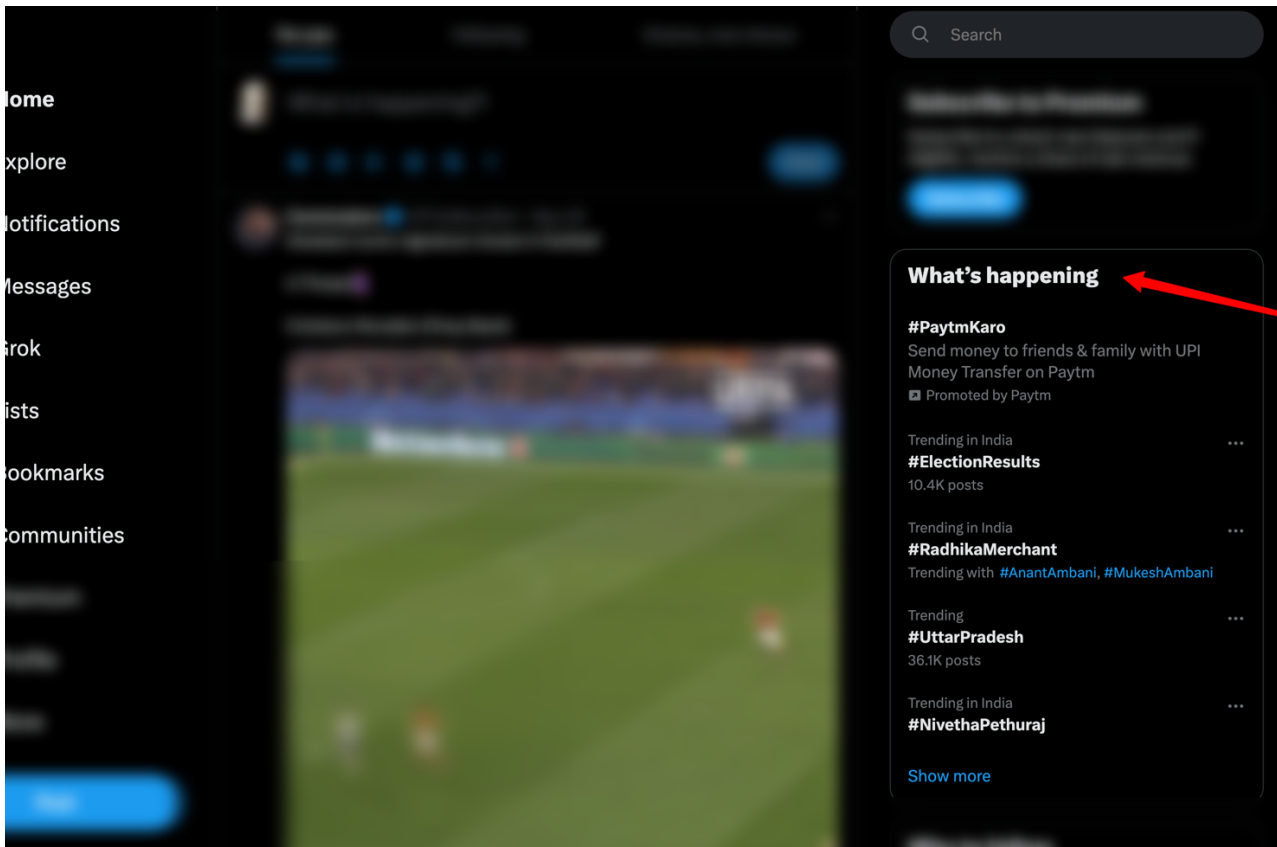
Our marketing tech product is still in stealth mode.

Your task

Web scraping with Selenium and ProxyMesh, storing the data in MongoDB, showing a list on a webpage. This task will test your ability to work with web automation tools, proxies, and data extraction techniques.

What you have to do:

1. Write a Selenium script that can read the Twitter home page (on your local computer) and fetch the top 5 trending topics under “What’s Happening” section from the homepage.



2. To access Twitter, create/use your own Twitter account, since log in required to see this page.
3. Use ProxyMesh such that each new request to scrape the trending topics is sent from a new IP address.
4. Create a unique ID for each time the Selenium script is run and store the results in a MongoDB database (fields required – unique ID, name of trend1, 2, 3, 4, 5, date and time of end of Selenium script, IP address used). We won't need anything else.
5. Create a simple HTML page which has a button, which when clicked, will run the Selenium script, and then show results in the following manner:

Expected result:

HTML page:

[Click here to run the script.](#)

When the link is click, it shows the following text:

These are the most happening topics as on {Date and Time of end of Selenium Script}

- Name of trend1
- Name of trend2
- Name of trend3
- Name of trend4
- Name of trend5

The IP address used for this query was XXX.XXX.XXX.XXX.

Here's a JSON extract of this record from the MongoDB:

```
[
  {
    "_id": { "XXX": "XXXXXXXX" },
    "nameoftrend1": "XXXX",
    "nameoftrend2": "XXXX",
    ....
  ]
```

[Click here to run the query again.](#)