

Machine Learning Engineer / Data Scientist / Data Science Intern

At CloudTara Technologies, we are a team of skilled problem-solvers with a deep technology background and want to drive companies to be technology and data-centric through the use of modern technology and actionable insights thereby bringing meaningful value, efficiency and competitive advantage to the business.

Job Description

CloudTara Technologies is hiring for Machine Learning Engineers / Data Scientists (for full-time applicants) and Data Science Interns (for part-time/interns) to help us businesses provide next-gen digital experience to our consumers by building revolutionary, cloud-based, scalable and reliable products using Agile-Scrum methodology.

Problem Statement

Build a web scraper that extracts the data of all cashless hospitals across India from the publicly available website of insurance companies, using Python, Selenium and other relevant tools.

For the given problem statement, the requirements are briefly as follows:

1. Build a web scraper bot that extracts pan-India data of cashless hospitals from the website of insurance companies mentioned in the Resources section.
2. Push the fetched data to a database of your choice. We prefer PostgreSQL or MySQL.
3. [OPTIONAL BONUS] Deploy the scraper service as a web API to trigger with a web request and allow the user to download the resultant data in form of a CSV file.

Resources

List of Insurance Companies [Pick Any 1]

- <https://www.adityabirlacapital.com/healthinsurance/#!/hospital-search>
- <https://general.bajajallianz.com/BagicNxt/hm/hmSearchExternalHosp.do>
- <https://www.careinsurance.com/health-plan-network-hospitals.html>
- <https://www.hdfcergo.com/locators/cashless-hospitals-network/>
- <https://www.starhealth.in/network-hospitals>



How to Submit Assignment

Keep the subject line of your email as “YOUR NAME - Assignment” (eg: “Shubhang Y - Assignment”).

Note:

- Please share the following information in the submission mailer:
 - Full-time: Current/Last Job Role, Company Name, Last drawn CTC, Notice Period / Joining Date Availability
 - Interns: Stream/branch of education, Current Semester, Joining Date Availability
- All submissions should include the project code (as a zip or hosted API link) along with a clear readme/setup file
- Other reference material

We are expecting the following attachments in your submissions:

1. Resume
2. Jupyter notebook (.ipynb)
3. The **cashless_COMPANYNAME.csv** file which will contain the scraped data from the insurance company's website. The records should be of all hospitals across India and should not contain blank lines in between.
4. Requirements.txt file
5. Text file containing sample input and expected response

Email your submissions to sadanand@cloudtara.com.

The last date for submission is June 12, 2021 (the date differs based on your batch, refer to the message which was sent to you). Feel free to reach out to us on the same email if you have any questions related to the assignment or submission date.

