

## ASSIGNMENT-1

### WEB SCRAPING

In all the following questions, you have to use BeautifulSoup to scrape different websites and collect data as per the requirement of the question.

Every answer to the question should be in form of a python function which should take URL as the parameter. Use Jupyter Notebooks to program, upload it on your GitHub and send the link of the Jupyter notebook to your SME.

- 1) Write a python program to display IMDB's Top rated 100 Indian movies' data <https://www.imdb.com/list/ls056092300/> (i.e. name, rating, year of release) and make data frame.
- 2) Write a python program to scrape product name, price and discounts from <https://peachmode.com/search?q=bags>
- 3) Write a python program to scrape cricket rankings from [icc-cricket.com](http://icc-cricket.com). You have to scrape:
  - a) Top 10 ODI teams in men's cricket along with the records for matches, points and rating.
  - b) Top 10 ODI Batsmen along with the records of their team and rating.
  - c) Top 10 ODI bowlers along with the records of their team and rating.
- 4) Write a python program to scrape details of all the posts from <https://www.patreon.com/coreyms>. Scrape the heading, date, content and the likes for the video from the link for the youtube video from the post.
- 5) Write a python program to scrape house details from mentioned URL. It should include house title, location, area, EMI and price from <https://www.nobroker.in/>. Enter three localities which are Indira Nagar, Jayanagar, Rajaji Nagar.
- 6) Write a python program to scrape first 10 product details which include product name, price, Image URL from <https://www.bewakoof.com/bestseller?sort=popular>.
- 7) Please visit <https://www.cnbc.com/world/?region=world> and scrap-
  - a) headings
  - b) date
  - c) News link
- 8) Please visit <https://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/most-downloaded-articles/> and scrap-
  - a) Paper title
  - b) date
  - c) Author