

# WEB SCRAPING – ASSIGNMENT 4

## Read all the problem statements, notes carefully and scrape the required data using any web scraping tool of your choice.

* **You have to handle commonly occurring EXCEPTIONS by using exception handling programing. To get information about selenium Exceptions. You may visit following links:**

## https://selenium-python.readthedocs.io/api.html

* 1. **https:**[**//www.guru99.com/exception-handling-selenium.html**](http://www.guru99.com/exception-handling-selenium.html)

## https://stackoverflow.com/questions/38022658/selenium-python-handling-no-such-element- exception/38023345

1. Scrape the details of most viewed videos on YouTube from Wikipedia. Url

= [https://en.wikipedia.org/wiki/List\_of\_most-viewed\_YouTube\_videos Y](https://en.wikipedia.org/wiki/List_of_most-viewed_YouTube_videos)ou need to find following details: A) Rank

* 1. Name
  2. Artist
  3. Upload date
  4. Views

1. Scrape the details team India’s international fixtures from bcci.tv. Url = https[://www.bcci.tv/.](http://www.bcci.tv/)

You need to find following details:

1. Series
2. Place
3. Date
4. Time

Note: - From bcci.tv home page you have reach to the international fixture page through code.

1. Scrape the details of State-wise GDP of India from statisticstime.com. Url = <http://statisticstimes.com/>

You have to find following details: A) Rank

1. State
2. GSDP(18-19)- at current prices
3. GSDP(19-20)- at current prices
4. Share(18-19)
5. GDP($ billion)

Note: - From statisticstimes home page you have to reach to economy page through code.

1. Scrape the details of trending repositories on Github.com. Url = https://github.com/

You have to find the following details:

1. Repository title
2. Repository description
3. Contributors count
4. Language used



Note: - From the home page you have to click on the trending option from Explore menu through code.

1. Scrape the details of top 100 songs on billiboard.com. Url = [https:/www.billboard.com/](https://www.billboard.com/) You have to find the following details:
2. Song name
3. Artist name
4. Last week rank
5. Peak rank
6. Weeks on board

Note: - From the home page you have to click on the charts option then hot 100-page link through code.

1. Scrape the details of Highest selling novels.
2. Book name
3. Author name
4. Volumes sold
5. Publisher
6. Genre

Url - <https://www.theguardian.com/news/datablog/2012/aug/09/best-selling-books-all-time-fifty-shades-grey-compare>

1. Scrape the details most watched tv series of all time from imdb.com.

Url = https:[//www.imdb.com/list/ls095964455/ Y](http://www.imdb.com/list/ls095964455/)ou have to find the following details:

1. Name
2. Year span
3. Genre
4. Run time
5. Ratings
6. Votes
7. Details of Datasets from UCI machine learning repositories.

Url = https://archive.ics.uci.edu/ You have to find the following details:

1. Dataset name
2. Data type
3. Task
4. Attribute type
5. No of instances
6. No of attribute G) Year

Note: - from the home page you have to go to the Show All Dataset page through code.