

# Assignment 1: Web Scraping & Data Analysis

Oct 07, 2023

In this assignment, you should work with data from <http://quotes.toscrape.com/>  
(Online Quotes)

## Quotes to Scrape

[Logout](#)

*"The world as we have created it is a process of our thinking. It cannot be changed without changing our thinking."*

by [Albert Einstein](#) (about) – (Goodreads page)

Tags: [change](#) [deep-thoughts](#) [thinking](#) [world](#)

*"It is our choices, Harry, that show what we truly are, far more than our abilities."*

by [J.K. Rowling](#) (about) – (Goodreads page)

Tags: [abilities](#) [choices](#)

*"There are only two ways to live your life. One is as though nothing is a miracle. The other is as though everything is a miracle."*

by [Albert Einstein](#) (about) – (Goodreads page)

Tags: [inspirational](#) [life](#) [live](#) [miracle](#) [miracles](#)

*"The person, be it gentleman or lady, who has not pleasure in a good novel, must be intolerably stupid."*

by [Jane Austen](#) (about) – (Goodreads page)

Tags: [aliteracy](#) [books](#) [classic](#) [humor](#)

### Top Ten tags

[love](#)  
[inspirational](#)  
[life](#)  
[humor](#)  
[books](#)  
[reading](#)  
[friendship](#)  
[friends](#)  
[truth](#)  
[simile](#)

Quotes to Scrape is a digital haven for those seeking a daily dose of inspiration, wit, and wisdom. This website is a treasure trove of thought-provoking and memorable quotes from a wide range of authors, thinkers, and personalities. Whether you're looking for words of wisdom to brighten your day or seeking quotes to add depth to your writing, Quotes to Scrape has you covered.

At Quotes to Scrape, you'll find an extensive collection of quotes that cover a multitude of topics, from love and life to success and motivation. Each quote is accompanied by the name of the author, allowing you to explore the perspectives of various individuals who have contributed to the world of literature and philosophy. In this project, we will use the requests and regular expressions we have learned before **to scrape all Quotes from the provided website.**

**Task1** You are required to **scrape 100 Quotes (all Quotes)** from the website and save result into 'your\_name+id.csv'. This file should contain data with the following columns: **(40 marks)**

Quotes Content.	5 marks
Tags	5 marks
Author	5 marks

Birthday of the author	5 marks
Country of the author	5 marks
Genre of the author	5 marks
Average rating of the author	5 marks
Number of reviews	5 marks

You are encouraged to explore data **with more properties** if needed.

**Task2.** You are required to do a data analysis on the data. What do you think is interesting about this data? Tell a story about some interesting thing you have discovered by looking at the data. **(60 marks)**

For example, which quotes you might read? Does the type of tags affect quotes you might read? Who is your favorite author?

**Note:** This is an open topic project. You are required to provide a novel topic and demonstrate your hypotheses (view points) with **data analysis and figures illustrations.**

The reports and running code (web scraping + data analysis) should be submitted using [Jupyter Notebook](#) file.

#### **Submission Checklist:**

Yes/No	Items
	Jupyter Notebook code
	your_name+id.csv

#### **Marking Guidelines**

Marking Criteria	
<b>Idea (5 marks)</b>	<ul style="list-style-type: none"> <li>• Presents a novel idea</li> <li>• Clearly demonstrate your viewpoints.</li> <li>• Demonstrates good understanding of the topic.</li> </ul>

<b>Discussion (30 marks)</b>	<ul style="list-style-type: none"> <li>• Provide convincing arguments to your viewpoints.</li> <li>• Backs up arguments with appropriate data analysis results.</li> <li>• Visualize data analysis results by using more than 5 figures.</li> </ul>
<b>Organization (20 marks)</b>	<ul style="list-style-type: none"> <li>• Use of figures to support ideas discussed in the report.</li> <li>• The quality of the figures.</li> <li>• These figures should be informative.</li> <li>• Use of sub-titles and/or clear topic sentences.</li> <li>• Use multiple visualization methods (line, bar, pie chart, etc, ).</li> </ul>
<b>Writing Style (5 marks)</b>	<ul style="list-style-type: none"> <li>• Concise writing style</li> <li>• Strong scientific writing without grammatical errors.</li> </ul>