

# WEB SCRAPING FINAL PROJECT

Sevinj Guluyeva

444423

s.guluyeva@student.uw.edu.pl

## 1. Introduction

EnGadget website is chosen to scrape points of different gadgets from different technology categories. Engadget defines itself as a blog network and covers consumer electronics.

For the final project, we are scraping engadget, critics and user scores for each product from different categories. Doing this may help to gain insights on popularity and usability for a consumer electronic. Further analysis can be done by user since we choose to provide a list of scores and a **dictionary** holding different products for each category.

## 2. Technical Details

As initial analysis, we could see that each product is assigned to a “StretchedBox” class. Parsing this lets us get all products in the page for scraping their content. Score boxes are hold on h3 tags. There are six h3 tags and every second tag corresponds to scores by Engadget score, critics score and user score, respectively.

We start our scrapers by first giving it a list of categories to scrape. For each category, scraper will get hrefs of products and send a request to them for further parsing. Scraper combine each score seperated by a comma (,) and this can be used to calculate sum of all Engadget scores in a category.

## 3. Scrapers' Output

Scrapers print output for scores of each product on each category. Later on, same output will provide sum of all engadgets for a category.

Example output from BeautifulSoup4 is:

```
=====laptops=====
```

Score for product dell-xps-13-plus:

- > Engadget Score:84
- > Critics Score:N/A
- > Users Score:N/A

Score for product lenovo-yoga-9i:

- > Engadget Score:90
- > Critics Score:N/A
- > Users Score:N/A

Score for product macbook-air-m2-2022:

- > Engadget Score:96
- > Critics Score:N/A
- > Users Score:N/A

Score for product asus-rog-zephyrus-duo-16:

- > Engadget Score:86
- > Critics Score:N/A

## WEB SCRAPING FINAL PROJECT

Sevinj Guluyeva

444423

s.guluyeva@student.uw.edu.pl

> Users Score:N/A

Score for product macbook-pro-13-inch-2022:

> Engadget Score:81

> Critics Score:N/A

> Users Score:N/A

Score for product surface-laptop-go-2:

> Engadget Score:86

> Critics Score:N/A

> Users Score:N/A

Score for product xps-15-2022:

> Engadget Score:93

> Critics Score:N/A

> Users Score:N/A

Score for product acer-predator-triton-500-se-2022:

> Engadget Score:84

> Critics Score:N/A

> Users Score:N/A

Score for product razer-blade-15-2022:

> Engadget Score:86

> Critics Score:N/A

> Users Score:N/A

Score for product asus-rog-flow-z13:

> Engadget Score:80

> Critics Score:N/A

> Users Score:N/A

Score for product zephyrus-g14-2022:

> Engadget Score:85

> Critics Score:N/A

> Users Score:N/A

Score for product x14:

> Engadget Score:85

> Critics Score:N/A

> Users Score:N/A

Category laptops Engadget Score: 1036

=====

=====phones=====

[....] Cut for brevity

# WEB SCRAPING FINAL PROJECT

Sevinj Guluyeva

444423

s.guluyeva@student.uw.edu.pl

=====

Further analysis can be done by user.

Performance for bs4 scraper as seconds is: 176.2908451

## 4. Data Analysis

Scraping results show that laptops and phones have more scores assigned to them. This seems to be a natural outcome of these categories being most used and bought tech gadgets by consumers. We may also see that external drives did not get any review by Engadget and hypothesize that as a result of consumers' reduced interest in buying them.

Further analysis can be done for users' and critics' scores to enlighten consumer electronics trends and help users decide which product covers their needs.

## 5. Participants

All three scrapers are developed by Sevinj Guluyeva. For development purposes, documents of respective modules were used and as a supplementary, different forums such as StackOverFlow was also used.