

## ShopGrok Python Exercises

Create a new public git repository and answer the following questions.

1. Create a python environment with the following installations. Please share the commands used to set-up the environment.
  - a. scrapy==2.4
  - b. shub
  - c. scrapy-crawlera
  - d. google-cloud-storage
  - e. scrapy-sessions
2. Write a basic python scrapy crawler using [scrapy library](#) for <https://tackleworldadelaide.com.au/> that:
  - a. **Navigates** through the top level-menu (ie Reels, Rods, etc) and **loops** through the submenu items within (ie Spinning Reels, Baitcast Reels, etc). This should be completed using xpath
  - b. Visits the category page and **loops** through each product tile on the page
  - c. **Yields** an item for each product with the following fields:
    - i. Sku\_name
    - ii. Image\_url
    - iii. Price\_now
    - iv. Price\_was
    - v. Product\_url
3. Write a basic python scrapy crawler using [scrapy library](#) that visits <https://www.surfboardempire.com.au/products.json?page=1> and:
  - a. **Loads** each product using the response JSON data
  - b. **Yields** an item for each product with the following fields:
    - i. Sku\_name
    - ii. Product\_id
    - iii. Brand
    - iv. Product\_url
  - c. **Loops** through each page until **all** products have been found
4. Write the python code that uses **regular expressions** to extract the numerical total number of products (ie 101) from the raw text below

```
<div>
    <span class="actionBar__text">    Showing 18 of 101
products.
    </span>
```

</div>