In [1]:

```
from bs4 import BeautifulSoup
import requests
import pandas as pd


# internshala has three pages with python/Django development category
# so by using format function we can scrap all pages.
# best ex. of the web with multiple pages
def get_url(page_no):
    url = 'https://internshala.com/internships/python%2Fdjango-internship/page-{}'.format(page_no)
    return url
def gather_information():
    data = {'Category':[],'company_name':[],'location':[],'start_date':[],'duration':[],'stipend':[],'apply_by':[],'c
    for i in range(1,4):
        res = requests.get(get_url(i))
        markup = res.content
        soup = BeautifulSoup(markup,'lxml')
        posts = soup.find_all('div',class_='container-fluid individual_internship')
        for i in posts:
            data['Category'].append(i.find('div',class_='heading_4_5').text)
            data['company_name'].append(i.find('a',class_='link_display_like_text').text.strip())
            data['location'].append(i.find('a',class_='location_link').text)
            data['start_date'].append(i.find('div',class_='other_detail_item').find_next(id='start-date-first').text.
            data['duration'].append(i.find('div',class_='item_body',id=False).text.strip())
            data['stipend'].append(i.find('span',class_='stipend').text)
            data['apply_by'].append(i.find('div',class_='apply_by').find('div',class_='item_body').text)
            data['detail'].append('https://internshala.com'+i.find('div',class_='button_container').a['href'])

    # create dataframe
    df = pd.DataFrame(data)
    # save file
    df.to_csv('internship.csv')

if __name__ == "__main__":
    gather_information()
```

In [ ]: