**Scraping of Car Website**

Assessment task given by company to scraping the given website [www.gotoauto.ca](https://github.com/sasikala07/Assessment_Scrapingtest/blob/main/www.gotoauto.ca)

and extract the car details i.e.,car name,car price and each car description from inventory pages and save it into csv file and description in text file within three days.

I scrapped the car details from the website using BeautifulSoup and saved the car\_name and car\_price in csv file format and saved each car description in one text file. I also tried to use selenium in task\_2,I saved one car description in text file format.

**About Tools**

1.**BeautifulSoup:** is a Python library for pulling data out of HTML and XML files. It works with your favourite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree. It transforms a complex HTML document into a tree of Python objects.

2.**Selenium**: Selenium Python bindings provide a simple API to write functional/acceptance tests using Selenium WebDriver like Firefox, Ie, Chrome, Remote etc.

**STEPS for Task**

**Task1:**

**Step1:** Firstly I import packages for scraping the website like,

BeautifulSoup,Urllib request library to download web html code.

*url\_req=Request('*[*https://www.gotoauto.ca/inventory/?pg=1',headers={'User-Agent':'Mozilla/5.0*](https://www.gotoauto.ca/inventory/?pg=1',headers=%7B'User-Agent':'Mozilla/5.0)*'})*

*url=urlopen(url\_req).read()*

**Step2**: list the number of webpages from range 1-3 for getting car list

*url\_list****=****['{}{}'****.****format(new\_url,str(page))* ***for*** *page* ***in*** *range(1,4)]*

*soup****=****BeautifulSoup(page,'html.parser')*

**Step3**: query the website and return the html to the variable 'page' and parse html

using Beautifulsoup also takes out the <div> of name and get its value.

*name****=****[x****.****get\_text(strip****=True****)* ***for*** *x* ***in*** *soup****.****find\_all('h4',{'class':'name desc\_l5'})]*

*price****=****[x****.****get\_text(strip****=True****)* ***for*** *x* ***in*** *soup****.****find\_all('span',{"class":'price'})]*

**Step4**: Remove unwanted characters and scrape the car name and MSRP (price) of

Those cars from the inventory pages and make data frames using pandas.

**Step5:** Save the scraping details into csv.file

**Task2:**

**Using BeautifulSoup**

**Step1:** Firstly I import packages for scraping the website like,

BeautifulSoup,Urllib request library to download web html code.

**Step2**: *using find\_all() method get all href values from tag <a> to get each car list*

list\_link**=**[str(link**.**get('href')) **for** link **in** soup**.**find\_all('a',class\_**=**'vehicle\_title\_link')]

**Step3**: Get the href links containing description of each car and scrap the data

url\_new**=**'<https://www.gotoauto.ca/inventory/?pg=1>'

url\_list**=**['{}{}'**.**format(url\_new,str(page)) **for** page **in** link]

**Step4**: *parse the html using beautifulsoup and takeout the descriptions from div tag*

soup**=**BeautifulSoup(page,'html.parser')

descs**=**[x**.**get\_text(strip**=True**) **for** x **in** soup**.**find\_all('div',class\_**=**'tab-pane fade show active')]

**Step5**: *save each car description from one page as text.file*

**with** open("task2.txt", "w") **as** output:

output**.**write(str(desc))

**Using Selenium**

**Step1**: Firstly install Selenium and webdriver (google chrome) pip install selenium=3.141.0 and import

packages . **from** selenium **import** webdriver

**Step2**: *create a new instance of google chrome to help open an url in google chrome and now acces google*

*chrome and open website that wants to be scraping*

driver**=**webdriver**.**Chrome('/home/sasikala/Downloads/chromedriver\_linux64/chromedriver')

driver**.**get('https://www.gotoauto.ca/inventory/')

**Step3**: Next ,*extract one car description on inventory page using Xpath function on selenium*

car\_desc**=**driver**.**find\_elements\_by\_xpath('//\*[@id="note"]/div/div[2]')[0]

desc**=**car\_desc**.**text

**Step4**: Save the description as text file

**with** open("task2\_selenium.txt", "w") **as** output:

output**.**write(str(car\_desc))

**Conclusion:**

I completed the task,Beutifulsoup and selenium are the most used scraping tools .I tried to complete task2 using selenium .Scraping the website I extract the car\_name and price of each page and get car descriptions of one page.