

# Forward School

**Program Code: J620-002-4:2020**

**Program Name: FRONT-END SOFTWARE DEVELOPMENT**

**Title : Case Study - Webscrapping on Trip Advisor**

**Name: Chong Mun Chen**

**IC Number: 960327-07-5097**

**Date : 4/7/2023**

**Introduction : Practising more with BeautifulSoup and Selenium on this Trip Advisor exercise.**

**Conclusion : I know more about the extensions and limitations of using the methods available in the BeautifulSoup and Selenium packages.**

Instruction: Use Selenium and Beautiful Soup to extract the first 5 comments and their titles from the trip advisor link.

Url: [https://www.tripadvisor.com/Airline\\_Review-d8729220-Reviews-AirAsia-AirAsia-Berhad-Malaysia](https://www.tripadvisor.com/Airline_Review-d8729220-Reviews-AirAsia-AirAsia-Berhad-Malaysia) ([https://www.tripadvisor.com/Airline\\_Review-d8729220-Reviews-AirAsia-AirAsia-Berhad-Malaysia](https://www.tripadvisor.com/Airline_Review-d8729220-Reviews-AirAsia-AirAsia-Berhad-Malaysia))

```
In [102]: ▶ from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.common.exceptions import ElementNotInteractableException, NoSuchElementException
from bs4 import BeautifulSoup

driver = webdriver.Chrome('C:\\Users\\ACER\\Desktop\\ChromeDriver\\chromedriver.exe')
```

```
In [103]: ▶ import time

driver.get("https://www.tripadvisor.com/Airline_Review-d8729220-Reviews-Air
time.sleep(5)

comments = []
for page in range(1, 6):
    soup = BeautifulSoup(driver.page_source, 'html.parser')
    soup_list = soup.find_all("div", attrs={"class": "WAlIg _T"})

    header = ''
    body = ''
    for items in range(len(soup_list)):
        header = soup_list[items].find("a", attrs={"class": "Qwuub"}).text
        body = soup_list[items].find('div', attrs={'class': 'vTVDC'}).find(

        if body.endswith("..."):
            body = body[:-1]

        comment = f"{header}:\n\t{body}"
        comments.append(comment)

    if page == 1:
        driver.find_element(By.XPATH, '//*[@id="component_1"]/div/div[5]/di
        time.sleep(5)
    elif page == 5:
        break
    else:
        driver.find_element(By.XPATH, '//*[@id="component_1"]/div/div[5]/di
        time.sleep(5)

driver.quit()

for i, comment in enumerate(comments):
    print(f"Comment {i+1}\n{comment}")
    print()
```

## Comment 1

No refund from Aiaasia for flight cancellation more than 8 months:

I booked Airaisa with booking no. T1ZDJV but cancelled without any reason or explanation, so I request to refund on 22/10/22, but 8 months passed till now still no refund, it kept saying being processed, just wondering is it just automatically replied or really being processed???

## Comment 2

Not good:

There is No food to speak of only the basics which you have to purchase as an extra. There is No alcoholic drinks. The seats are very narrow and I only have a 32 inch waste. Our connecting Flight from Phuket to Kuala Lumpa was changed from 7.20pm evening to 8.10 in the morning of the same day and we were only given 8 hours notice. They do not have a call centre in any country and the only way you can try to contact them is through there website and they do not answer you still waiting for answers to our emails. Instead of having an hour stop over we had 12 hours plus it took a whole day off our holiday. The young lady at the

Reference: <https://medium.com/ymedialabs-innovation/web-scraping-using-beautiful-soup-and-selenium-for-dynamic-page-2f8ad15efe25> (<https://medium.com/ymedialabs-innovation/web-scraping-using-beautiful-soup-and-selenium-for-dynamic-page-2f8ad15efe25>)