|  |  |
| --- | --- |
| A blue logo with a black background  Description automatically generated | **AIR UNIVERSITY** |
| **DEPARTMENT OF COMPUTER SCIENCE** |
| **Lab Task 6** |

**Student Name: Hamza Umer Farooq Reg. No: 200789**

**Subject: Data Science Semester: VIII**

**Objective: Web Scraping**

**ASSESSMENT:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Attributes** | **Excellent**  **(5)** | **Good**  **(4)** | **Average**  **(3)** | **Satisfactory**  **(2)** | **Unsatisfactory (1)** |
| **Ability to Conduct**  Task |  |  |  |  |  |
| **Ability to assimilate the results** |  |  |  |  |  |
| **Effective use of theorems/postulates/formulas** |  |  |  |  |  |

Total Marks:

Obtained Marks:

**REPORT ASSESSMENT:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Attributes** | **Excellent**  **(5)** | **Good**  **(4)** | **Average**  **(3)** | **Satisfactory**  **(2)** | **Unsatisfactory**  **(1)** |
| **Data presentation** |  |  |  |  |  |
| **Experimental results** |  |  |  |  |  |
| **Conclusion** |  |  |  |  |  |

from bs4 import BeautifulSoup as bs

import requests

link = 'https://www.rottentomatoes.com/m/kung\_fu\_panda\_4/reviews'

page = requests.get(link)

page

page.content

soup= bs(page.content)

soup.prettify()

names = soup.find\_all('a',class\_='display-name')

names

cust\_name = []

for i in range(0,len(names)):

    cust\_name.append(names[i].get\_text().replace("  ", "").replace("\n", ""))

cust\_name

len(cust\_name)

text = soup.find\_all('p',class\_='review-text')

text

reviews = []

for i in range(0,len(text)):

    reviews.append(text[i].get\_text())

reviews

len(reviews)

reviews[:] = [titles.lstrip('\n') for titles in reviews]

reviews

reviews[:] = [texts.rstrip('\n') for texts in reviews]

reviews

rating = soup.find\_all('score-icon-critic-deprecated')

rating

rate = []

for i in range(0,len(rating)):

    if rating[i].get("state") == "rotten":

      rate.append("Negative")

    else:

      rate.append("Positive")

rate

len(rate)

import pandas as pd

df1 = pd.DataFrame()

df2 = pd.DataFrame()

df3 = pd.DataFrame()

df1['Customer Name']=cust\_name

df2['Reviews']=reviews

df3['Ratings']=rate

result = pd.concat([df1,df2,df3],axis=1)

result.head()

result.shape

result.to\_csv('reviews.csv',index=True)

Screenshots:  
