

VI Semester PBL Report

WEB SCRAPING USING PYTHON PROJECT

BY

L. Sushmitha	20641A6701
P. Koushik	20641A6706
A. Nandhini	20641A6711
D. Shiva Shankar	20641A6721
B. Tharun	21645A6701

Under the Guidance of

Mrs. Vemulamma (Assistant professor)



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
(DATA SCIENCE)

VAAGDEVI COLLEGE OF ENGINEERING

(UGC Autonomous, Accredited by NBA, Accredited by NAAC with “A”)

Bollikunta, Khila Warangal (Mandal), Warangal Urban-506 005

VAAGDEVI COLLEGE OF ENGINEERING

(UGC Autonomous, Accredited by NBA, Accredited by NAAC with “A”)

Bollikunta, Khila Warangal (Mandal), Warangal Urban-506 005(T.S)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(DATA SCIENCE)



CERTIFICATE

This is to certify that the project work carried out by **L.Sushmitha(20641A6701), P.Koushik(20641A6706), A.Nandhini(20641A6711), D.Shiva Shankar(20641A6721), B.Tharun(21645A6701)**, is considered as part of the Project Based Learning in the 3rd semester of B.Tech Computer Science and Engineering (Data Science) during the academic year 2022-23.

Guide

Head of the Department

External Examiner

ACKNOWLEDGEMENT

The development of the project in Project Based learning (B20DS13,6th Semester) though was an arduous task, it has been made by the help of many people. We are pleased to express our thanks to the people whose suggestions, comments, criticisms greatly encouraged us in betterment of the project.

We would like to express my sincere gratitude and indebtedness to my project Guide

P. VEMULAMMA for his/her valuable suggestions and interest throughout the course of this project.

We would like to express my sincere thanks and profound gratitude to **Dr. K. PRAKASH**, principal of Vaagdevi College of Engineering, for his support, guidance and encouragement in the course of our project.

We are also thankful to the Head of the Department **Dr. AYESHA BANU**, for providing excellent infrastructure and a nice atmosphere for completing this project successfully.

We are highly thankful to the Project Coordinators for their valuable suggestions, encouragement and motivations for completing this project successfully.

We are thankful to all other faculty members for their encouragement We convey my heartfelt thanks to the lab staff for allowing me to use the required equipment whenever needed.

We are also thankful to P. VEMULAMMA and supporting staff of Vaagdevi college for their valuable suggestions, encouragement and motivations for completing this project successfully.

Finally, we would like to take this opportunity to thank my family for their support through the work. I sincerely acknowledge and thank all those who gave directly or indirectly their support in completion of this work.

L. Sushmitha	(20641A6701)
P. Koushik	(20641A6706)
A. Nandhini	(20641A6711)
D. Shiva Shankar	(20641A6721)
B. Tharun	(21645A6701)

DECLARATION

We declare that the work reported in the project entitled "ADMINSTRATION MODULE" is a record of work done by us in partial fulfilment for the award for the degree Bachelor of Technology in computer science of engineering, Vaagdevi college of engineering affiliated by JNTUH ,Accredited by NBA ,under the guide of P. Vemulamma CSE Department .

we hereby declare that the project work bears no resemblance to any other project submitted at vaagdevi college of engineering or any other college/university for the award of the degree.

Signature:.....

P. Vemulamma

Internal Guide

ABSTRACT:

- Internet is the vastest information and data source ever built by mankind. However, it is a huge collection of heterogeneous and poorly structured data, difficult to collect in a manual way and complicated to use in automated processes.
- Over the last years, techniques and tools have surged, allowing data collection and conversion to structured data to be managed by B2C and B2B systems. This article offers an introduction to web scraping techniques and some of the most popular and novel techniques for data extraction and reuse in complex processes.
- The possibilities to take benefit of such data are many, including areas like Open Government Data, Big Data, Business Intelligence, aggregators and comparators, development of new applications and mashups, among others.

Contents

ABSTRACT.....	5
INTRODUCTION	7
ALGORITHM.....	12
FLOW CHART	13
IMPLEMENTATION	14
CONCLUSION.....	28
SOFTWARE AND HARDWARE PACKAGE SPECIFICATIONS	29

INTRODUCTION:

- Every single day, several petabytes of information are published via the Internet in various formats, such as HTML, PDF, CSV or XML. Curiously, HTML is not necessarily the most common format to publish content.
- For instance, 70% of the content indexed by Google is extracted from PDF documents. This is an additional obstacle for the different roles involved in data extraction from multiple sources.
- Journalists, researchers, data analysts, sales agents or software developers are some examples of professionals typically using the copy-and-paste technique to get information in a specific format and export it to a spreadsheet, an executive report or to some data exchange format such as XML, JSON or any of the several vocabularies available based on them.
- Focusing on data acquisition from HTML (i.e, a web document), this article explains the mechanisms and tools that may help us to minimize the tedious duties of data extraction from the Internet. With regards to the commitment of transparency and data openness that public administrations have assumed over the last decade, scraping and crawling are techniques that may be useful.
- Whereas current information systems in use by public administrations consider these techniques, a relevant amount of web sites, content management systems and ECMs (Enterprise Content Management) do not.
- Exchanging these systems for new ones imply considerable economic efforts. Scraping tools provide an alternative to exchanging which minimizes such efforts. Open Government Data and Transparency Policies should take advantage of this opportunity.

WHAT IS WEB SCRAPING ?

- The dictionary meaning of word 'Scraping' implies getting something from the web. Here two questions arise: What we can get from the web and How to get that. The answer to the first question is 'data'. Data is indispensable for any programmer and the basic requirement of every programming project is the large amount of useful data.
- The answer to the second question is a bit tricky, because there are lots of ways to get data. In general, we may get data from a database or data file and other sources. But what if we need large amount of data that is available online? One way to get such kind of data is to manually search (clicking away in a web browser) and save (copy-pasting into a spreadsheet or file) the required data. This method is quite tedious and time consuming.
- Another way to get such data is using web scraping. Web scraping, also called web data mining or web harvesting, is the process of constructing an agent which can extract, parse, download and organize useful information from the web automatically.
- In other words, we can say that instead of manually saving the data from websites, the web scraping software will automatically load and extract data from multiple websites as per our requirement.

COMPONENTS OF WEB SCRAPER

A web scraper consists of the following components: Web Crawler Module A very necessary component of web scraper, web crawler module, is used to navigate the target website by making HTTP or HTTPS request to the URLs. The crawler downloads the unstructured data (HTML contents) and passes it to extractor, the next module.

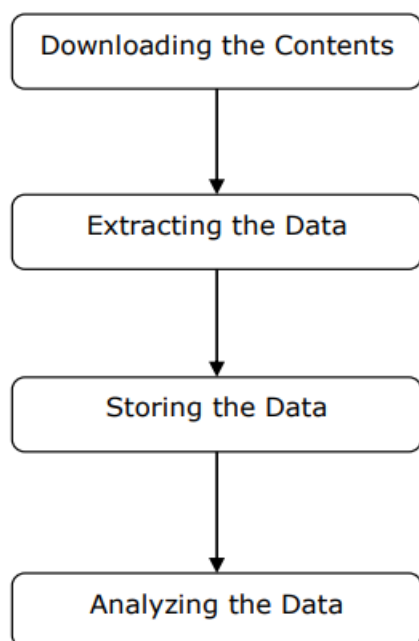
Extractor : The extractor processes the fetched HTML content and extracts the data into semistructured format. This is also called as a parser module and uses different parsing techniques like Regular expression, HTML Parsing, DOM parsing or Artificial Intelligence for its functioning.

Data Transformation and Cleaning Module : The data extracted above is not suitable for ready use. It must pass through some cleaning module so that we can use it. The methods like

String manipulation or regular expression can be used for this purpose. Note that extraction and transformation can be performed in a single step also.

Storage Module : After extracting the data, we need to store it as per our requirement. The storage module will output the data in a standard format that can be stored in a database or JSON or CSV format.

WORKING OF WEB SCRAPER



Python Web Scraping – Getting Started with Python

WHY PYTHON FOR WEB SCRAPING ?

- Python is a popular tool for implementing web scraping. Python programming language is also used for other useful projects related to cyber security, penetration testing as well as digital forensic applications. Using the base programming of Python, web scraping can be performed without using any other third party tool. Python programming language is gaining huge popularity and the reasons that make Python a good fit for web scraping projects are as below:

- **Syntax Simplicity** : Python has the simplest structure when compared to other programming languages. This feature of Python makes the testing easier and a developer can focus more on programming.
- **Inbuilt Modules** : Another reason for using Python for web scraping is the inbuilt as well as external useful libraries it possesses. We can perform many implementations related to web scraping by using Python as the base for programming.
- **Open Source** : Programming Language Python has huge support from the community because it is an open source programming language. Wide range of Applications Python can be used for various programming tasks ranging from small shell scripts to enterprise web applications.

Python Modules for Web Scraping

- Web scraping is the process of constructing an agent which can extract, parse, download and organize useful information from the web automatically. In other words, instead of manually saving the data from websites, the web scraping software will automatically load and extract data from multiple websites as per our requirement.

Requests:

- It is a simple python web scraping library. It is an efficient HTTP library used for accessing web pages. With the help of Requests, we can get the raw HTML of web pages which can then be parsed for retrieving the data. Before using requests, let us understand its installation. Installing Requests We can install it in either on our virtual environment or on the global installation. With the help of pip command.

```
PS C:\Users\banot\OneDrive\Desktop\web> pip3 install requests
Defaulting to user installation because normal site-packages is not writeable
Collecting requests
  Downloading requests-2.31.0-py3-none-any.whl (62 kB)
    |████████████████████| 62 kB 335 kB/s
Collecting charset-normalizer<4,>=2
  Downloading charset-normalizer-3.1.0-cp310-cp310-win_amd64.whl (97 kB)
    |████████████████████| 97 kB 1.6 MB/s
```

Beautifulsoup :

we are using BeautifulSoup for parsing that HTML data.

Beautiful Soup is a Python library for pulling data out of HTML and XML files. It works with your favorite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree. It commonly saves programmers hours or days of work.

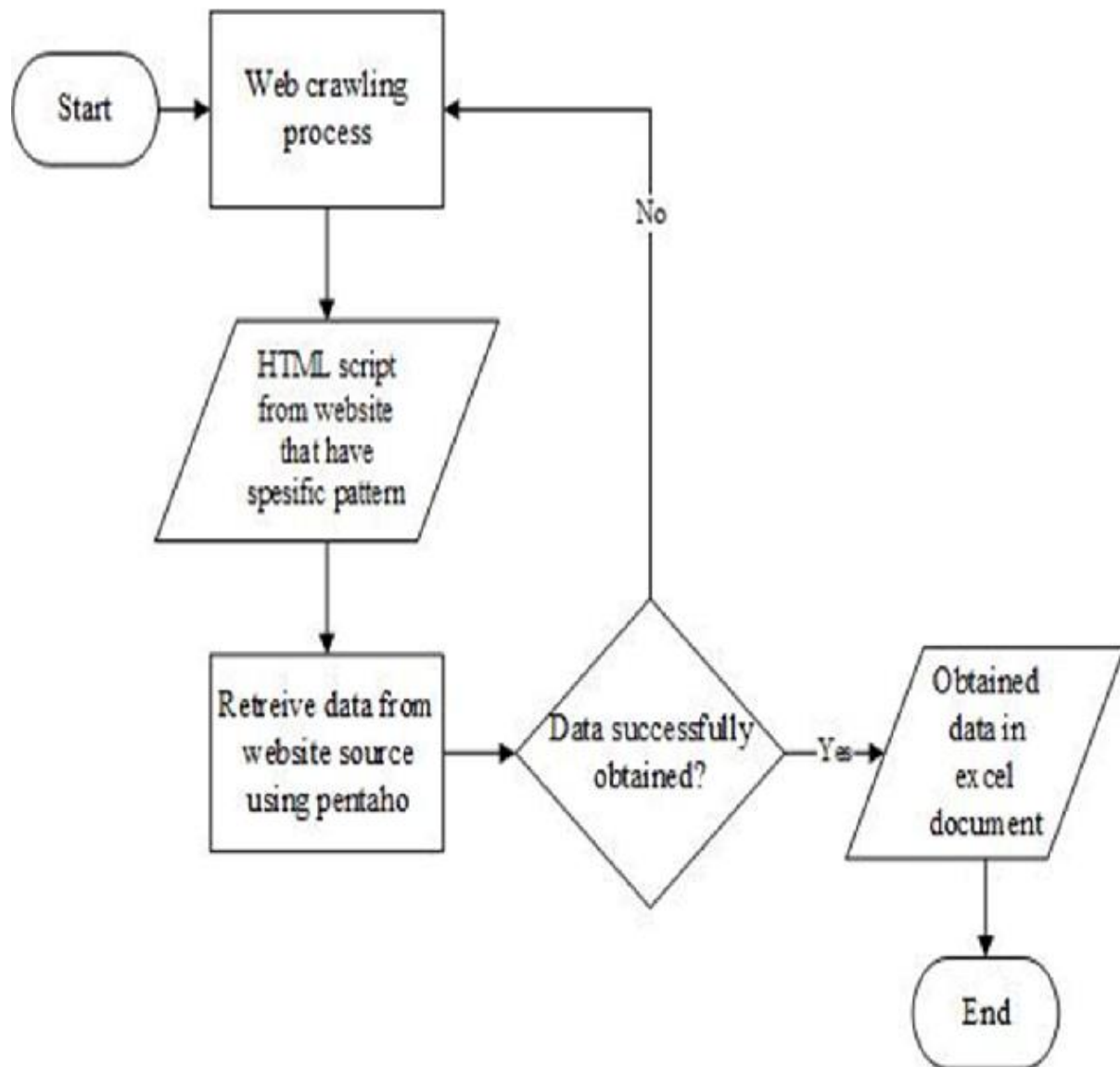
```
PS C:\Users\banot\OneDrive\Desktop\web> pip3 install bs4
Defaulting to user installation because normal site-packages is not writeable
Collecting bs4
  Downloading bs4-0.0.1.tar.gz (1.1 kB)
Collecting beautifulsoup4
  Downloading beautifulsoup4-4.12.2-py3-none-any.whl (142 kB)
    |████████████████████| 142 kB 1.3 MB/s
Collecting soupsieve>1.2
  Downloading soupsieve-2.4.1-py3-none-any.whl (36 kB)
Using legacy 'setup.py install' for bs4, since package 'wheel' is not installed.
Installing collected packages: soupsieve, beautifulsoup4, bs4
  Running setup.py install for bs4 ... done
Successfully installed beautifulsoup4-4.12.2 bs4-0.0.1 soupsieve-2.4.1
```

ALGORITHM :

We can understand the working of a web scraper in simple steps as shown in the diagram given above.

- **Step 1:** Downloading Contents from Web Pages In this step, a web scraper will download the requested contents from multiple web pages.
- **Step 2:** Extracting Data The data on websites is HTML and mostly unstructured. Hence, in this step, web scraper will parse and extract structured data from the downloaded contents.
- **Step 3:** Storing the Data Here, a web scraper will store and save the extracted data in any of the format like CSV, JSON or in database.
- **Step 4:** Analyzing the Data After all these steps are successfully done, the web scraper will analyze the data thus obtained.

FLOW CHART :



IMPLEMENTATION:

INPUT :

```
#importing libraries
from bs4 import BeautifulSoup
import requests, openpyxl

excel = openpyxl.Workbook()
print(excel.sheetnames)
sheet = excel.active
sheet.title = 'top rated movies'
print(excel.sheetnames)
sheet.append([' Movie ', ' movie Name ', ' Year Of Release ', ' IMDB Rating '])
try:
#Requesting the webserver to access the data
    source = requests.get('https://www.imdb.com/chart/top/')
    source.raise_for_status()

    soup = BeautifulSoup(source.text, 'html.parser')

    movies = soup.find('tbody', class_='lister-list').find_all('tr')

    for movie in movies:

        name = movie.find('td', class_='titleColumn').a.text
```

```
rank = movie.find('td',  
class_='titleColumn').get_text(strip=True).split(',')[0]  
year = movie.find('td', class_='titleColumn').span.text.strip('()')  
rating = movie.find('td', class_='ratingColumn  
imdbRating').strong.text  
print(rank,name,year,rating)  
sheet.append([rank,name,year,rating])
```

```
except Exception as e:
```

```
    print(e)
```

```
excel.save('IMDB Movie Rating.xlsx')
```

OUTPUT :

```
PS C:\Users\banot\OneDrive\Desktop\web> & "C:/Program Files/Python310/python.exe"  
c:/Users/banot/OneDrive/Desktop/web.py
```

```
['Sheet']
```

```
['top rated movies']
```

- 1.The Shawshank Redemption(1994) The Shawshank Redemption 1994 9.2
- 2.The Godfather(1972) The Godfather 1972 9.2
- 3.The Dark Knight(2008) The Dark Knight 2008 9.0
- 4.The Godfather Part II(1974) The Godfather Part II 1974 9.0
- 5.12 Angry Men(1957) 12 Angry Men 1957 9.0
- 6.Schindler's List(1993) Schindler's List 1993 8.9
- 7.The Lord of the Rings: The Return of the King(2003) The Lord of the Rings: The Return of the King 2003 8.9
- 8.Pulp Fiction(1994) Pulp Fiction 1994 8.8
- 9.The Lord of the Rings: The Fellowship of the Ring(2001) The Lord of the Rings: The Fellowship of the Ring 2001 8.8
- 10.Il buono Il buono, il brutto, il cattivo 1966 8.8
- 11.Spider-Man: Across the Spider-Verse(2023) Spider-Man: Across the Spider-Verse 2023 8.8
- 12.Forrest Gump(1994) Forrest Gump 1994 8.8
- 13.Fight Club(1999) Fight Club 1999 8.7
- 14.The Lord of the Rings: The Two Towers(2002) The Lord of the Rings: The Two Towers 2002 8.7
- 15.Inception(2010) Inception 2010 8.7

- 16.Star Wars: Episode V - The Empire Strikes Back(1980) Star Wars: Episode V - The Empire Strikes Back 1980 8.7
- 17.The Matrix(1999) The Matrix 1999 8.7
- 18.GoodFellas(1990) GoodFellas 1990 8.7
- 19.One Flew Over the Cuckoo's Nest(1975) One Flew Over the Cuckoo's Nest 1975 8.6
- 20.Se7en(1995) Se7en 1995 8.6
- 21.It's a Wonderful Life(1946) It's a Wonderful Life 1946 8.6
- 22.Shichinin no samurai(1954) Shichinin no samurai 1954 8.6
- 23.The Silence of the Lambs(1991) The Silence of the Lambs 1991 8.6
- 24.Saving Private Ryan(1998) Saving Private Ryan 1998 8.6
- 25.Cidade de Deus(2002) Cidade de Deus 2002 8.6
- 26.Interstellar(2014) Interstellar 2014 8.6
- 27.La vita è bella(1997) La vita è bella 1997 8.6
- 28.The Green Mile(1999) The Green Mile 1999 8.6
- 29.Star Wars(1977) Star Wars 1977 8.5
- 30.Terminator 2: Judgment Day(1991) Terminator 2: Judgment Day 1991 8.5
- 31.Back to the Future(1985) Back to the Future 1985 8.5
- 32.Sen to Chihiro no kamikakushi(2001) Sen to Chihiro no kamikakushi 2001 8.5
- 33.The Pianist(2002) The Pianist 2002 8.5
- 34.Psycho(1960) Psycho 1960 8.5
- 35.Gisaengchung(2019) Gisaengchung 2019 8.5
- 36.Léon(1994) Léon 1994 8.5
- 37.The Lion King(1994) The Lion King 1994 8.5
- 38.Gladiator(2000) Gladiator 2000 8.5
- 39.American History X(1998) American History X 1998 8.5

- 40.The Departed(2006) The Departed 2006 8.5
- 41.Whiplash(2014) Whiplash 2014 8.5
- 42.The Prestige(2006) The Prestige 2006 8.5
- 43.The Usual Suspects(1995) The Usual Suspects 1995 8.5
- 44.Casablanca(1942) Casablanca 1942 8.5
- 45.Hotaru no haka(1988) Hotaru no haka 1988 8.5
- 46.Seppuku(1962) Seppuku 1962 8.5
- 47.The Intouchables(2011) The Intouchables 2011 8.5
- 48.Modern Times(1936) Modern Times 1936 8.4
- 49.Once Upon a Time in the West(1968) Once Upon a Time in the West 1968 8.4
- 50.Nuovo Cinema Paradiso(1988) Nuovo Cinema Paradiso 1988 8.4
- 51.Rear Window(1954) Rear Window 1954 8.4
- 52.Alien(1979) Alien 1979 8.4
- 53.City Lights(1931) City Lights 1931 8.4
- 54.Apocalypse Now(1979) Apocalypse Now 1979 8.4
- 55.Memento(2000) Memento 2000 8.4
- 56.Django Unchained(2012) Django Unchained 2012 8.4
- 57.Raiders of the Lost Ark(1981) Raiders of the Lost Ark 1981 8.4
- 58.WALL·E(2008) WALL·E 2008 8.4
- 59.The Lives of Others(2006) The Lives of Others 2006 8.4
- 60.Sunset Blvd.(1950) Sunset Blvd. 1950 8.4
- 61.Paths of Glory(1957) Paths of Glory 1957 8.4
- 62.Avengers: Infinity War(2018) Avengers: Infinity War 2018 8.4
- 63.The Shining(1980) The Shining 1980 8.4

- 64.The Great Dictator(1940) The Great Dictator 1940 8.4
- 65.Witness for the Prosecution(1957) Witness for the Prosecution 1957 8.4
- 66.Spider-Man: Into the Spider-Verse(2018) Spider-Man: Into the Spider-Verse 2018 8.4
- 67.Aliens(1986) Aliens 1986 8.3
- 68.American Beauty(1999) American Beauty 1999 8.3
- 69.The Dark Knight Rises(2012) The Dark Knight Rises 2012 8.3
- 70.Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb(1964) Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb 1964 8.3
- 71.Inglourious Basterds(2009) Inglourious Basterds 2009 8.3
- 72.Oldeuboi(2003) Oldeuboi 2003 8.3
- 73.Coco(2017) Coco 2017 8.3
- 74.Amadeus(1984) Amadeus 1984 8.3
- 75.Toy Story(1995) Toy Story 1995 8.3
- 76.Braveheart(1995) Braveheart 1995 8.3
- 77.Das Boot(1981) Das Boot 1981 8.3
- 78.Joker(2019) Joker 2019 8.3
- 79.Avengers: Endgame(2019) Avengers: Endgame 2019 8.3
- 80.Mononoke-hime(1997) Mononoke-hime 1997 8.3
- 81.Good Will Hunting(1997) Good Will Hunting 1997 8.3
- 82.Once Upon a Time in America(1984) Once Upon a Time in America 1984 8.3
- 83.Kimi no na wa.(2016) Kimi no na wa. 2016 8.3
- 84.3 Idiots(2009) 3 Idiots 2009 8.3
- 85.Singin' in the Rain(1952) Singin' in the Rain 1952 8.3
- 86.Tengoku to jigoku(1963) Tengoku to jigoku 1963 8.3
- 87.Requiem for a Dream(2000) Requiem for a Dream 2000 8.3

- 88.Capharnaüm(2018) Capharnaüm 2018 8.3
- 89.Toy Story 3(2010) Toy Story 3 2010 8.3
- 90.Star Wars: Episode VI - Return of the Jedi(1983) Star Wars: Episode VI - Return of the Jedi 1983 8.3
- 91.Idi i smotri(1985) Idi i smotri 1985 8.3
- 92.Eternal Sunshine of the Spotless Mind(2004) Eternal Sunshine of the Spotless Mind 2004 8.3
- 93.2001: A Space Odyssey(1968) 2001: A Space Odyssey 1968 8.3
- 94.Jagten(2012) Jagten 2012 8.3
- 95.Reservoir Dogs(1992) Reservoir Dogs 1992 8.3
- 96.Lawrence of Arabia(1962) Lawrence of Arabia 1962 8.2
- 97.Citizen Kane(1941) Citizen Kane 1941 8.2
- 98.M - Eine Stadt sucht einen Mörder(1931) M - Eine Stadt sucht einen Mörder 1931 8.2
- 99.Ikiru(1952) Ikiru 1952 8.2
- 100.North by Northwest(1959) North by Northwest 1959 8.2
- 101.The Apartment(1960) The Apartment 1960 8.2
- 102.Vertigo(1958) Vertigo 1958 8.2
- 103.Le fabuleux destin d'Amélie Poulain(2001) Le fabuleux destin d'Amélie Poulain 2001 8.2
- 104.Double Indemnity(1944) Double Indemnity 1944 8.2
- 105.A Clockwork Orange(1971) A Clockwork Orange 1971 8.2
- 106.Full Metal Jacket(1987) Full Metal Jacket 1987 8.2
- 107.Hamilton(2020) Hamilton 2020 8.2
- 108.Scarface(1983) Scarface 1983 8.2
- 109.Incendies(2010) Incendies 2010 8.2

- 110.Heat(1995) Heat 1995 8.2
- 111.Up(2009) Up 2009 8.2
- 112.To Kill a Mockingbird(1962) To Kill a Mockingbird 1962 8.2
- 113.The Sting(1973) The Sting 1973 8.2
- 114.Jodaeiye Nader az Simin(2011) Jodaeiye Nader az Simin 2011 8.2
- 115.Metropolis(1927) Metropolis 1927 8.2
- 116.L.A. Confidential(1997) L.A. Confidential 1997 8.2
- 117.Die Hard(1988) Die Hard 1988 8.2
- 118.Indiana Jones and the Last Crusade(1989) Indiana Jones and the Last Crusade 1989 8.2
- 119.Snatch(2000) Snatch 2000 8.2
- 120.Ladri di biciclette(1948) Ladri di biciclette 1948 8.2
- 121.Taxi Driver(1976) Taxi Driver 1976 8.2
- 122.Taare Zameen Par(2007) Taare Zameen Par 2007 8.2
- 123.1917(2019) 1917 2019 8.2
- 124.Dangal(2016) Dangal 2016 8.2
- 125.Der Untergang(2004) Der Untergang 2004 8.2
- 126.Top Gun: Maverick(2022) Top Gun: Maverick 2022 8.2
- 127.Per qualche dollaro in più(1965) Per qualche dollaro in più 1965 8.2
- 128.Batman Begins(2005) Batman Begins 2005 8.2
- 129.The Kid(1921) The Kid 1921 8.2
- 130.Some Like It Hot(1959) Some Like It Hot 1959 8.2
- 131.The Father(2020) The Father 2020 8.2
- 132.The Wolf of Wall Street(2013) The Wolf of Wall Street 2013 8.2
- 133.All About Eve(1950) All About Eve 1950 8.2

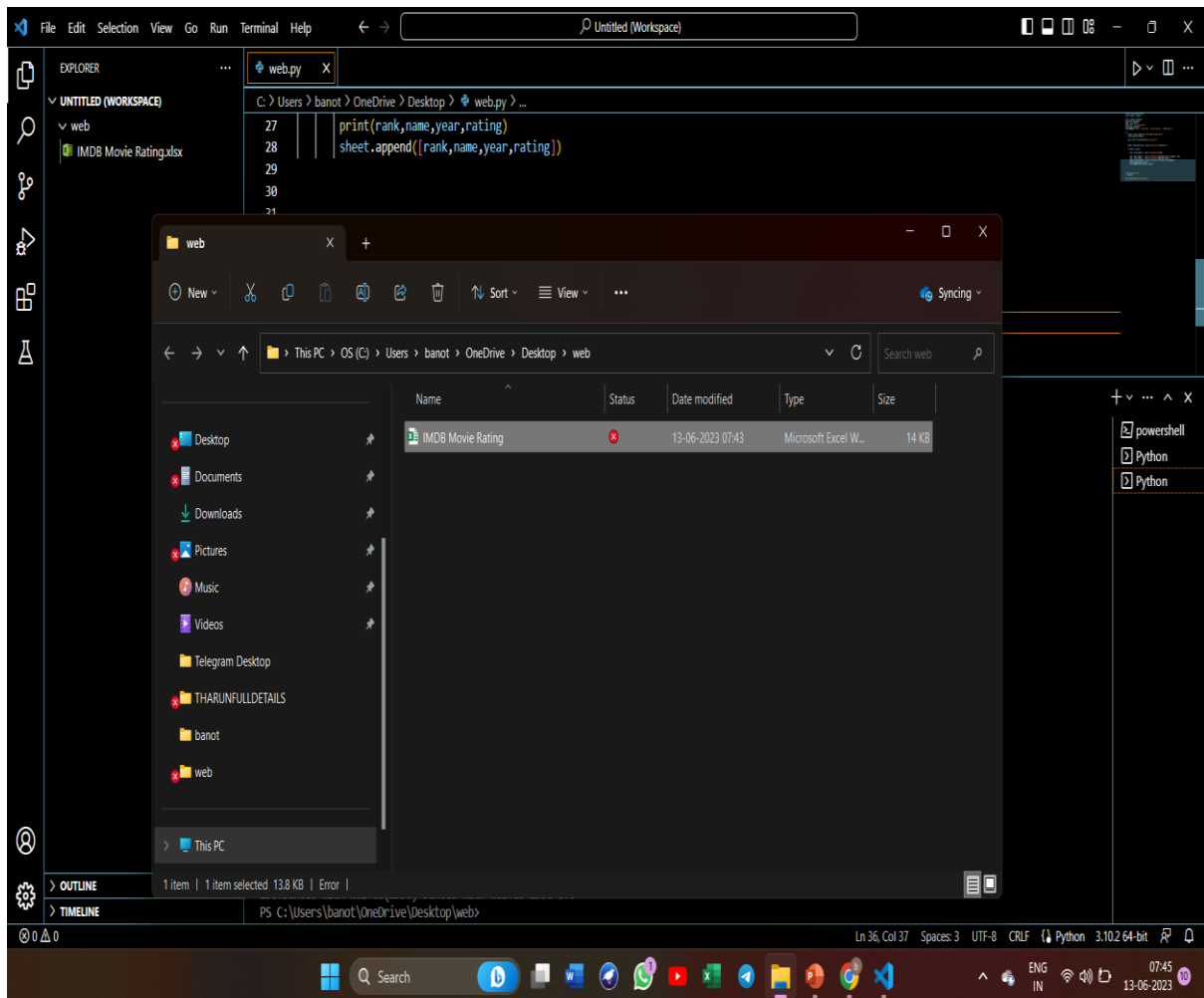
- 134.Green Book(2018) Green Book 2018 8.2
- 135.Judgment at Nuremberg(1961) Judgment at Nuremberg 1961 8.2
- 136.Ran(1985) Ran 1985 8.2
- 137.The Truman Show(1998) The Truman Show 1998 8.2
- 138.Casino(1995) Casino 1995 8.2
- 139.There Will Be Blood(2007) There Will Be Blood 2007 8.2
- 140.Pan's Labyrinth(2006) Pan's Labyrinth 2006 8.2
- 141.Unforgiven(1992) Unforgiven 1992 8.2
- 142.Shutter Island(2010) Shutter Island 2010 8.2
- 143.The Sixth Sense(1999) The Sixth Sense 1999 8.2
- 144.Jurassic Park(1993) Jurassic Park 1993 8.2
- 145.A Beautiful Mind(2001) A Beautiful Mind 2001 8.2
- 146.The Treasure of the Sierra Madre(1948) The Treasure of the Sierra Madre 1948 8.1
- 147.Yôjinbô(1961) Yôjinbô 1961 8.1
- 148.Monty Python and the Holy Grail(1975) Monty Python and the Holy Grail 1975 8.1
- 149.No Country for Old Men(2007) No Country for Old Men 2007 8.1
- 150.Kill Bill: Vol. 1(2003) Kill Bill: Vol. 1 2003 8.1
- 151.The Great Escape(1963) The Great Escape 1963 8.1
- 152.Rashômon(1950) Rashômon 1950 8.1
- 153.The Thing(1982) The Thing 1982 8.1
- 154.Finding Nemo(2003) Finding Nemo 2003 8.1
- 155.Spider-Man: No Way Home(2021) Spider-Man: No Way Home 2021 8.1
- 156.The Elephant Man(1980) The Elephant Man 1980 8.1
- 157.Chinatown(1974) Chinatown 1974 8.1

- 158.Raging Bull(1980) Raging Bull 1980 8.1
- 159.V for Vendetta(2005) V for Vendetta 2005 8.1
- 160.Gone with the Wind(1939) Gone with the Wind 1939 8.1
- 161.Dial M for Murder(1954) Dial M for Murder 1954 8.1
- 162.Hauru no ugoku shiro(2004) Hauru no ugoku shiro 2004 8.1
- 163.Inside Out(2015) Inside Out 2015 8.1
- 164.Lock Lock, Stock and Two Smoking Barrels 1998 8.1
- 165.El secreto de sus ojos(2009) El secreto de sus ojos 2009 8.1
- 166.Three Billboards Outside Ebbing Three Billboards Outside Ebbing, Missouri 2017 8.1
- 167.Prisoners(2013) Prisoners 2013 8.1
- 168.Trainspotting(1996) Trainspotting 1996 8.1
- 169.The Bridge on the River Kwai(1957) The Bridge on the River Kwai 1957 8.1
- 170.Guardians of the Galaxy Vol. 3(2023) Guardians of the Galaxy Vol. 3 2023 8.1
- 171.Fargo(1996) Fargo 1996 8.1
- 172.Warrior(2011) Warrior 2011 8.1
- 173.Gran Torino(2008) Gran Torino 2008 8.1
- 174.Catch Me If You Can(2002) Catch Me If You Can 2002 8.1
- 175.Tonari no Totoro(1988) Tonari no Totoro 1988 8.1
- 176.Million Dollar Baby(2004) Million Dollar Baby 2004 8.1
- 177.Bacheha-Ye aseman(1997) Bacheha-Ye aseman 1997 8.1
- 178.The Gold Rush(1925) The Gold Rush 1925 8.1
- 179.Klaus(2019) Klaus 2019 8.1
- 180.Blade Runner(1982) Blade Runner 1982 8.1
- 181.Harry Potter and the Deathly Hallows: Part 2(2011) Harry Potter and the Deathly Hallows: Part 2 2011 8.1

- 182.12 Years a Slave(2013) 12 Years a Slave 2013 8.1
- 183.Before Sunrise(1995) Before Sunrise 1995 8.1
- 184.The Grand Budapest Hotel(2014) The Grand Budapest Hotel 2014 8.1
- 185.Ben-Hur(1959) Ben-Hur 1959 8.1
- 186.Gone Girl(2014) Gone Girl 2014 8.1
- 187.On the Waterfront(1954) On the Waterfront 1954 8.1
- 188.Barry Lyndon(1975) Barry Lyndon 1975 8.1
- 189.The General(1926) The General 1926 8.1
- 190.Smultronstället(1957) Smultronstället 1957 8.1
- 191.In the Name of the Father(1993) In the Name of the Father 1993 8.1
- 192.The Deer Hunter(1978) The Deer Hunter 1978 8.1
- 193.Hacksaw Ridge(2016) Hacksaw Ridge 2016 8.1
- 194.The Third Man(1949) The Third Man 1949 8.1
- 195.Le salaire de la peur(1953) Le salaire de la peur 1953 8.1
- 196.Salinui chueok(2003) Salinui chueok 2003 8.1
- 197.Sherlock Jr.(1924) Sherlock Jr. 1924 8.1
- 198.Relatos salvajes(2014) Relatos salvajes 2014 8.1
- 199.Mr. Smith Goes to Washington(1939) Mr. Smith Goes to Washington 1939 8.1
- 200.Mad Max: Fury Road(2015) Mad Max: Fury Road 2015 8.1
- 201.Mary and Max.(2009) Mary and Max. 2009 8.1
- 202.Dead Poets Society(1989) Dead Poets Society 1989 8.1
- 203.How to Train Your Dragon(2010) How to Train Your Dragon 2010 8.1
- 204.Monsters Monsters, Inc. 2001 8.1
- 205.Det sjunde inseglet(1957) Det sjunde inseglet 1957 8.1

- 206.Jaws(1975) Jaws 1975 8.1
- 207.Room(2015) Room 2015 8.1
- 208.The Big Lebowski(1998) The Big Lebowski 1998 8.1
- 209.Tôkyô monogatari(1953) Tôkyô monogatari 1953 8.1
- 210.Ford v Ferrari(2019) Ford v Ferrari 2019 8.1
- 211.La passion de Jeanne d'Arc(1928) La passion de Jeanne d'Arc 1928 8.0
- 239.Amores perros(2000) Amores perros 2000 8.0
- 240.Rebecca(1940) Rebecca 1940 8.0
- 241.Cool Hand Luke(1967) Cool Hand Luke 1967 8.0
- 242.The Sound of Music(1965) The Sound of Music 1965 8.0
- 243.It Happened One Night(1934) It Happened One Night 1934 8.0
- 244.Les quatre cents coups(1959) Les quatre cents coups 1959 8.0
- 245.Persona(1966) Persona 1966 8.0
- 246.The Help(2011) The Help 2011 8.0
- 247.The Iron Giant(1999) The Iron Giant 1999 8.0
- 248.Life of Brian(1979) Life of Brian 1979 8.0
- 249.Aladdin(1992) Aladdin 1992 8.0
- 250.Dances with Wolves(1990) Dances with Wolves 1990 8.0

Then the excel sheet is created with name '**IMDB Movie Rating.xlsx**'



By open the excel file, the data is shown as below,

The screenshot shows an Excel spreadsheet titled "IMDB Movie Rating - Saved". The data is organized in a table with the following columns: Rank, Movie Name, Year, and IMDB Rating. The table lists 27 movies, with the top-rated being "The Shawshank Redemption" (Rank 1, Rating 9.2).

Rank	Movie Name	Year	IMDB Rating
1	The Shawshank Redemption	1994	9.2
2	The Godfather	1972	9.2
3	The Dark Knight	2008	9.0
4	The Godfather Part II	1974	9.0
5	12 Angry Men	1957	9.0
6	Schindler's List	1993	8.9
7	The Lord of the Rings: The Fellowship of the Ring	2003	8.9
8	Pulp Fiction	1994	8.8
9	The Lord of the Rings: The Two Towers	2001	8.8
10	Il buono, il brutto, il cattivo	1966	8.8
11	Spider-Man	2002	8.8
12	Forrest Gump	1994	8.8
13	Fight Club	1999	8.7
14	The Lord of the Rings: The Return of the King	2003	8.7
15	Inception	2010	8.7
16	Star Wars: The Force Awakens	2015	8.7
17	The Matrix	1999	8.7
18	Goodfellas	1990	8.7
19	One Flew Over the Cuckoo's Nest	1975	8.6
20	Se7en	1995	8.6
21	It's a Wonderful Life	1946	8.6
22	Shichinin no Shikoku	1954	8.6
23	The Silence of the Lambs	1991	8.6
24	Saving Private Ryan	1998	8.6
25	Cidade de Deus	2002	8.6
26	Interstellar	2014	8.6
27	The Godfather Part I	1972	8.6

CONCLUSION :

- Web scraping is a familiar term that has gained importance because of the need to “free” data stored in PDF documents or web pages. Many professionals and researchers need the data in order to process it, analyse it and extract meaningful results.
- On the other hand, people dealing with B2B use cases need to access data from multiple sources to integrate it in new applications that provide added value and innovation.
- The market demands holistic web scraping solutions, that encompass cloud storage and ease to build interoperable APIs. In this article we have analysed and compared 3 web scraping solution: beautyfulsoup and requests.
- These solutions differ in their implementation details but share more common ground than believed: a visual picker, JSON results, data caching, background robots to gather data, programmatic APIs, etc.
- An experienced team, using these solutions, can develop services such as content aggregators, ranking tools, mobile apps, data monitoring systems, reputation management applications, business intelligence solutions, Big Data, etc.

HARDWARE AND SOFTWARE PACKAGE SPECIFICATIONS

HARDWARE NECESSITIES

Hardware choice is essential to the standard and potency of any software package. In Hardware choice, size and power necessities are necessary. Customer isolation will be with success run on the system with AN i3 processor with a minimum of four GB RAM and disc drive with 500GB and fifteen.6 inches to observe system performance. (Printer is needed for text output).

- Pentium processor ----- intel i3
- RAM capability ----- 8 GB
- Hard Disk ----- 256 GB

SOFTWARE NECESSITIES

- Operating System :: Windows 10
- Software: python
- Code editor: Visual Studio Code
- Python modules : requests module, beautifulsoup module of python
- Databases::Excel sheets ,pdf