

Ch-8 Web Programming Using Python(17 Marks)

web scraping using python(BeautifulSoup & requests)

- 9 Marks first code(Online Code)
- DeepLearning => 6 M theory code

```
In [2]: import requests
import bs4
url = "https://subslikescript.com/movie/Titanic-120338"
```

```
In [3]: source1 = requests.get(url)
print(source1)
```

<Response [200]>

```
In [4]: source = requests.get(url).text
print(source)
```

```
<body>
  <div class="page-wrapper">
    <header>
      <div id="logo">
        <a href="/" title="Subs like Script - al
1 Movies and TV Shows Transcripts">
          <span>SUBSLIKESCRIPT</span>
        </a>
      </div>
      <nav>
        <a target="_self" href="/movies" title
="All movies transcripts">Movies</a>
        <a target="_self" href="/series" title
="All Tv Shows transcripts">TV Shows</a>
      </nav>
    </header>
    <div class="main-wrapper">
      <main class="mainpage">
        <nav>
          <ul class="breadcumb">
```

```
In [5]: soup = bs4.BeautifulSoup(source, "html.parser")
print(soup)
```

```
zh.css" rel="stylesheet"/><script src="https://subscribescript.com/build/
assets/app-Cj664GvU.js" type="module"></script> <script async="" crossor
igin="anonymous" src="https://pagead2.googlesyndication.com/pagead/js/ad
sbygoogle.js?client=ca-pub-9736463211397065"></script>
</link></head>
<body>
<div class="page-wrapper">
<header>
<div id="logo">
<a href="/" title="Subs like Script - all Movies and TV Shows Transcript
s">
<span>SUBSLIKESCRIPT</span>
</a>
</div>
<nav>
<a href="/movies" target="_self" title="All movies transcripts">Movies</
a>
<a href="/series" target="_self" title="All Tv Shows transcripts">TV Sho
ws</a>
</nav>
```

```
In [6]: print(soup.prettify())
```

```
<!DOCTYPE html>
<html dir="ltr" lang="en">
<head>
  <!-- Global site tag (gtag.js) - Google Analytics -->
  <script async="" src="https://www.googletagmanager.com/gtag/js?id=UA-1
20598793-1">
  </script>
  <script>
    window.dataLayer = window.dataLayer || [];
    function gtag(){dataLayer.push(arguments);}
    gtag('js', new Date());

    gtag('config', 'UA-120598793-1');
  </script>
  <meta charset="utf-8"/>
  <title>
    Titanic (1997) Movie Script
  | Subs like Script
  </title>
  <script>
```

```
In [7]: title = soup.find('h1').text
print(title)
```

Titanic (1997) - full transcript

```
In [8]: transcript = soup.find('p').text
print(transcript) # first p prefer.
```

A seventeen-year-old aristocrat falls in love with a kind but poor artist aboard the luxurious, ill-fated R.M.S. Titanic.

```
In [9]: transcript = soup.find('p',class_="cue-line").text  
print(transcript) # then only one line print because p tag is end.
```

13 meters. You should see it.

```
In [10]: transcript = soup.find('div',class_="full-script").text  
print(transcript) # full script mate parent tag no class Levano.
```

13 meters. You should see it.

Okay, take her up and over the bow rail.

Mir 2, we're going over the bow.
Stay with us.

Okay, quiet. We're rolling.

Seeing her coming out of the
darkness like a ghost ship...

... ..

```
In [11]: title = soup.find('h1').get_text()  
print(title)
```

Titanic (1997) - full transcript

```
In [12]: transcript = soup.find('div',class_="full-script").get_text()  
print(transcript)
```

13 meters. You should see it.

Okay, take her up and over the bow rail.

Mir 2, we're going over the bow.
Stay with us.

Okay, quiet. We're rolling.

Seeing her coming out of the
darkness like a ghost ship...

... ..

```
In [13]: f = open("Titanic.txt", "w", encoding="utf-8")
         f.write(transcript)
```

Out[13]: 67373

Task

```
In [14]: import requests
         import bs4
         url = "https://www.politifact.com/factchecks/"
```

```
In [15]: source1 = requests.get(url)
         source = requests.get(url).text
         soup = bs4.BeautifulSoup(source, "html.parser")
         transcript = soup.find('a', class_="m-statement__name").text
         print(transcript)
```

X posts

- find(string)=>text,get_text()
- find_all(list)

```
In [16]: # 1 mcq fix
         # transcript = soup.find_all('a', class_="m-statement__name").text
         # print(transcript)
```

```
In [17]: dic = {"statement":[], "source":[], "date":[]}
```

```
In [18]: transcript = soup.find_all('a',class_="m-statement__name")  
print(transcript)
```

```
[<a class="m-statement__name" href="/personalities/tweets/" title="X posts">
    X posts
</a>, <a class="m-statement__name" href="/personalities/tweets/" title="X posts">
    X posts
</a>, <a class="m-statement__name" href="/personalities/donald-trump/" title="Donald Trump">
    Donald Trump
</a>, <a class="m-statement__name" href="/personalities/tweets/" title="X posts">
    X posts
</a>, <a class="m-statement__name" href="/personalities/social-media/" title="Social Media">
    Social Media
</a>, <a class="m-statement__name" href="/personalities/gavin-newsom/" title="Gavin Newsom">
    Gavin Newsom
</a>, <a class="m-statement__name" href="/personalities/tiktok-posts/" title="TikTok posts">
    TikTok posts
</a>, <a class="m-statement__name" href="/personalities/karoline-leavitt/" title="Karoline Leavitt">
    Karoline Leavitt
</a>, <a class="m-statement__name" href="/personalities/robert-reich/" title="Robert Reich">
    Robert Reich
</a>, <a class="m-statement__name" href="/personalities/tweets/" title="X posts">
    X posts
</a>, <a class="m-statement__name" href="/personalities/scott-bessent/" title="Scott Bessent">
    Scott Bessent
</a>, <a class="m-statement__name" href="/personalities/robert-f-kennedy-jr/" title="Robert F. Kennedy Jr.">
    Robert F. Kennedy Jr.
</a>, <a class="m-statement__name" href="/personalities/bernie-sanders/" title="Bernie Sanders">
    Bernie Sanders
</a>, <a class="m-statement__name" href="/personalities/gavin-newsom/" title="Gavin Newsom">
    Gavin Newsom
</a>, <a class="m-statement__name" href="/personalities/donald-trump/" title="Donald Trump">
    Donald Trump
</a>, <a class="m-statement__name" href="/personalities/alex-jones/" title="Alex Jones">
    Alex Jones
</a>, <a class="m-statement__name" href="/personalities/marcia-morey/" title="Marcia Morey">
    Marcia Morey
</a>, <a class="m-statement__name" href="/personalities/stephen-miller/" title="Stephen Miller">
    Stephen Miller
</a>, <a class="m-statement__name" href="/personalities/donald-trump/" title="Donald Trump">
    Donald Trump
</a>, <a class="m-statement__name" href="/personalities/donald-trump/" title="Donald Trump">
    Donald Trump
</a>, <a class="m-statement__name" href="/personalities/
```

```
alex-jones/" title="Alex Jones">
    Alex Jones
</a>, <a class="m-statement__name" href="/personalities/
social-media/" title="Social Media">
    Social Media
</a>, <a class="m-statement__name" href="/personalities/
hakeem-jeffries/" title="Hakeem Jeffries">
    Hakeem Jeffries
</a>, <a class="m-statement__name" href="/personalities/
pam-bondi/" title="Pam Bondi">
    Pam Bondi
</a>, <a class="m-statement__name" href="/personalities/
charles-blow/" title="Charles Blow">
    Charles Blow
</a>, <a class="m-statement__name" href="/personalities/
tweets/" title="X posts">
    X posts
</a>, <a class="m-statement__name" href="/personalities/
tweets/" title="X posts">
    X posts
</a>, <a class="m-statement__name" href="/personalities/
donald-trump/" title="Donald Trump">
    Donald Trump
</a>, <a class="m-statement__name" href="/personalities/
donald-trump/" title="Donald Trump">
    Donald Trump
</a>, <a class="m-statement__name" href="/personalities/
jared-moskowitz/" title="Jared Moskowitz">
    Jared Moskowitz
</a>]
```

```
In [19]: transcript = soup.find_all('a',class_="m-statement__name")
for i in transcript:
    b = i.text.strip()
    dic["source"].append(b)
```

```
In [20]: st = soup.find_all('div',class_="m-statement__quote")
print(st)
```

```
[<div class="m-statement__quote">
<a href="/factchecks/2025/may/23/tweets/LaMonica-McIver-CNN-altered-vide
o-singing/">
```

Rep. LaMonica McIver, D-N.J., started singing during a May 20 CNN interview.

```
</a>
</div>, <div class="m-statement__quote">
<a href="/factchecks/2025/may/23/tweets/no-a-shooting-outside-the-capita
l-jewish-museum-wa/">
```

The slaying of two Israeli embassy employees in Washington, D.C., is a false flag.

```
</a>
</div>, <div class="m-statement__quote">
<a href="/factchecks/2025/may/22/donald-trump/medicaid-waste-fraud-abuse
-changes/">
```

```
In [21]: st = soup.find_all('div',class_="m-statement__quote")
         for i in st:
             b = i.text.strip()
             dic["statement"].append(b)
```



```
In [22]: ft = soup.find_all('footer',class_="m-statement__footer")  
print(ft)
```

```
[<footer class="m-statement__footer">
  By Loreben Tuquero • May 23, 2025
</footer>, <footer class="m-statement__footer">
  By Maria Briceño • May 23, 2025
</footer>, <footer class="m-statement__footer">
  By Louis Jacobson • May 22, 2025
</footer>, <footer class="m-statement__footer">
  By Loreben Tuquero • May 22, 2025
</footer>, <footer class="m-statement__footer">
  By Grace Abels • May 21, 2025
</footer>, <footer class="m-statement__footer">
  By Amy Sherman • May 21, 2025
</footer>, <footer class="m-statement__footer">
  By Maria Briceño • May 21, 2025
</footer>, <footer class="m-statement__footer">
  By Louis Jacobson • May 21, 2025
</footer>, <footer class="m-statement__footer">
  By Loreben Tuquero • May 21, 2025
</footer>, <footer class="m-statement__footer">
  By Grace Abels • May 20, 2025
</footer>, <footer class="m-statement__footer">
  By Louis Jacobson • May 20, 2025
</footer>, <footer class="m-statement__footer">
  By Madison Czopek • May 16, 2025
</footer>, <footer class="m-statement__footer">
  By Louis Jacobson • May 16, 2025
</footer>, <footer class="m-statement__footer">
  By Amy Sherman • May 16, 2025
</footer>, <footer class="m-statement__footer">
  By Louis Jacobson • May 16, 2025
</footer>, <footer class="m-statement__footer">
  By Maria Briceño • May 16, 2025
</footer>, <footer class="m-statement__footer">
  By Paul Specht • May 15, 2025
</footer>, <footer class="m-statement__footer">
  By Maria Ramirez Uribe • May 15, 2025
</footer>, <footer class="m-statement__footer">
  By Louis Jacobson • May 14, 2025
</footer>, <footer class="m-statement__footer">
  By Amy Sherman • May 13, 2025
</footer>, <footer class="m-statement__footer">
  By Maria Briceño • May 13, 2025
</footer>, <footer class="m-statement__footer">
  By Maria Ramirez Uribe • May 12, 2025
</footer>, <footer class="m-statement__footer">
  By Louis Jacobson • May 12, 2025
</footer>, <footer class="m-statement__footer">
  By Maria Ramirez Uribe • May 9, 2025
</footer>, <footer class="m-statement__footer">
  By Loreben Tuquero • May 9, 2025
</footer>, <footer class="m-statement__footer">
  By Madison Czopek • May 9, 2025
</footer>, <footer class="m-statement__footer">
  By Grace Abels • May 9, 2025
</footer>, <footer class="m-statement__footer">
  By Madison Czopek • May 8, 2025
</footer>, <footer class="m-statement__footer">
  By Maria Ramirez Uribe • May 8, 2025
</footer>, <footer class="m-statement__footer">
```

```
In [23]: ft = soup.find_all('footer',class_="m-statement__footer")
for i in ft:
#     b = i.text.strip().split("•")[1]
    b = i.text[-31:-1].strip()
    dic["date"].append(b)
```

In [24]: `print(dic)`

```
{'statement': ['Rep. LaMonica McIver, D-N.J., started singing during a May 20 CNN interview.', 'The slaying of two Israeli embassy employees in Washington, D.C., is a false flag.', 'In the House bill, “we’re not changing Medicaid,” only cutting “waste, fraud and abuse.”', 'The Sean “Diddy” Combs trial testimony revealed that former President Barack Obama was “caught in secret meetings with Diddy’s drug runner.”', 'Video shows the Qatari emir saying, “I truly regret inviting Trump, the robber, to the Middle East.”', 'Que California ofrezca acceso asequible a cuidado de salud para los inmigrantes ilegalmente en EE.UU.', 'La presidenta de México Claudia Sheinbaum dijo que el choque del buque Cuauhtémoc contra el puente de Brooklyn en Nueva York no fue un accidente y puede ser un acto hostil.', 'President Donald Trump’s tax and spending bill “does not add to the deficit.”', 'While serving in Congress, Transportation Secretary Sean Duffy “voted against upgrading air traffic control systems.”', 'Former President Joe Biden’s cancer is a form of “turbo cancer” caused by mRNA vaccines.', '"Gasoline prices have collapsed under President Trump."', 'It’s “all true” that the measles vaccine wanes quickly, was never fully safety tested and contains fetal debris.', '"Republicans just unveiled their “big, beautiful bill,” which will take Medicaid & health insurance away from 13.7 million Americans."', 'On California offering affordable health care access for immigrants in the U.S. illegally.', 'Declining cargo traffic at U.S. ports “means we lose less money. ... When you say it's slowed down, that's a good thing.”', 'Un video muestra al presidente francés, Emmanuel Macron, al primer ministro del Reino Unido, Keir Starmer, y al canciller alemán, Friedrich Merz, con una bolsa de cocaína tras visitar al presidente ucraniano, Volodymyr Zelenskyy.', '"Google caved in to Trump. They erased Black History Month from the calendar."', 'The Immigration and Nationality Act “stripped Article III courts, that's the judicial branch, of jurisdiction over immigration cases.”', 'Under a new executive order, prescription drug prices will be reduced “almost immediately.”', 'The Digital Equity Act is a handout “based on race,” and “illegal.”', 'A video shows French President Emmanuel Macron, United Kingdom Prime Minister Keir Starmer and German Chancellor Friedrich Merz with a bag of cocaine after visiting Ukrainian President Volodymyr Zelenskyy.', '"In 1996, Congress specifically authorized the executive branch to conduct non-judicial deportations NOT SUBJECT TO DUE PROCESS."', '"About 20% of households with veterans rely upon” the Supplemental Nutrition Assistance Program.', 'En los primeros 100 días del presidente Donald Trump, las incautaciones de fentanilo salvaron entre “119 millones” y “258 millones” de vidas.', '"China makes 80% of all toys” and “90% of all Christmas goods” sold in the U.S.', 'Headline said, “Pope Leo XIV was a founding member of the 2020 anarchist Portland Autonomous Zone known as CHAZ.”', 'Robert Prevost, the Roman Catholic cardinal who was chosen to be the new pope, is a “registered Republican.”', '"We have now close to $10 trillion” in investments. “We’re talking about essentially two months.”', '"If people come into our country illegally, there’s a different standard” for due process.', 'Canadian tourism to Florida has declined by 80%.'], 'source': ['X posts', 'X posts', 'Donald Trump', 'X posts', 'Social Media', 'Gavin Newsom', 'TikTok posts', 'Karoline Leavitt', 'Robert Reich', 'X posts', 'Scott Bessent', 'Robert F. Kennedy Jr.', 'Bernie Sanders', 'Gavin Newsom', 'Donald Trump', 'Alex Jones', 'Marcia Morey', 'Stephen Miller', 'Donald Trump', 'Donald Trump', 'Alex Jones', 'Social Media', 'Hakeem Jeffries', 'Pam Bondi', 'Charles Blow', 'X posts', 'X posts', 'Donald Trump', 'Donald Trump', 'Jared Moskowitz'], 'date': ['May 23, 2025', 'May 23, 2025', 'May 22, 2025', 'May 22, 2025', 'May 21, 2025', 'May 21, 2025', 'May 21, 2025', 'May 21, 2025', 'May 21, 2025', 'May 20, 2025', 'May 20, 2025', 'May 16, 2025', 'May 16, 2025', 'May 16, 2025', 'May 16, 2025', 'May 15, 2025', 'May 15, 2025', 'May 14, 2025', 'May 13, 2025', 'May 13, 2025', 'May 12, 2025', 'May 12, 2025', 'May 9, 2025', 'May 9, 2025', 'May 9, 2025', 'May 9, 2025', 'May 8, 2025', 'May 8, 2025', 'May 7, 2025']}
```

PB : 354

```
In [28]: import bs4
f = open("Today 52 Week Low BSE_NSE Stocks Companies List - Ticker.html", encoding="utf-8")
soup = bs4.BeautifulSoup(f, "html.parser")
print(soup.prettify())
```

```
<!DOCTYPE html>
<html>
  <head>
    <!-- Global site tag (gtag.js) - Google Analytics -->
    <script async="" src="https://www.googletagmanager.com/gtag/js?id=UA-136614031-6">
    </script>
    <script>
      window.dataLayer = window.dataLayer || [];
      function gtag() { dataLayer.push(arguments); }
      gtag('js', new Date());
      gtag('config', 'UA-136614031-6');
    </script>
    <title>
      Today 52 Week Low BSE/NSE Stocks Companies List - Ticker
    </title>
    <meta charset="utf-8"/>
    <meta content="width=device-width, initial-scale=1, shrink-to-fit=no"
name="viewport"/>
```

```
In [56]: dic = {"sno.":[], "Company":[], "Price Rs : ":[], "Day low Rs.":[]}
```

Type Markdown and LaTeX: α^2

```
In [57]: com = soup.find_all('td', class_="left")
for i in com:
    b = i.text.strip()
    dic["Company"].append(b)
```

```
In [58]: num = soup.find_all('td', class_="Number")
for i in num:
    b = i.text.strip()
    dic["Price Rs : "].append(b)
```

```
In [59]: low = soup.find_all('span', class_="Number")
# sn=0
for i in low:
    b = i.text.strip()
    # sn+=1
    dic["Day low Rs."].append(b)
    # dic["sno."].append(sn)
```

In [60]: `print(dic)`

```
{'sno.': [], 'Company': ['Rachana Infrastru', 'Tirupati Forge', 'Sera Inve
stments&Fin', 'Viaz Tyres', 'Sellwin Traders', 'Soni Medicare', 'Sicagen I
ndia', 'Aspira Pathlab&Diagn', 'Patspin India', 'AG Universal', 'Arihant F
oundn. &Hsg', 'Global Offshore Serv', 'GTN Textiles', 'GTN Inds', 'Vivanza
Biosciences', 'TECIL Chem & Hydro', 'Milgrey Fin.&Invest', 'Zodiac-JRD-MK
J', 'Kanungo Financiers', 'CIL Nova Petro', 'Integ.Pro', 'Elango Inds', 'V
oltaire Leasing'], 'Price Rs : ': ['125.70', '7.80', '13.01', '46.35', '1
3.00', '19.22', '24.35', '24.85', '14.10', '42.10', '39.65', '7.85', '19.5
0', '11.95', '8.33', '21.60', '16.29', '31.95', '5.35', '16.05', '8.81',
'5.55', '11.99'], 'Day low Rs.': ['113.80', '7.75', '13.01', '45.00', '12.
01', '17.40', '24.00', '24.80', '14.00', '42.00', '38.50', '7.75', '19.5
0', '11.50', '8.33', '20.70', '14.81', '31.50', '4.85', '15.25', '8.81',
'5.55', '11.54']}
```

In [50]: `sr = soup.find_all("tr")`
`print(sr)`

```
[<tr>
<th class="left" data-sortable="true" data-width="5%">S.No.</th>
<th class="left" data-sortable="true">Company</th>
<th data-sortable="true" data-sorter="changeSorter">price <span class="m
uted">Rs.</span></th>
<th data-sortable="true" data-sorter="changeSorter">Day Low <span class
="muted">Rs.</span></th>
</tr>, <tr>
<td>1</td>
<td class="left"><a href="/company/RILINFRA">Rachana Infrastru</a></td>
<td class="Number">125.70</td>
<td><span class="Number">113.80</span></td>
</tr>, <tr>
<td>2</td>
<td class="left"><a href="/company/TIRUPATIFL">Tirupati Forge</a></td>
<td class="Number">7.80</td>
<td><span class="Number">7.75</span></td>
</tr>, <tr>
<td>3</td>
<td class="left"><a href="/company/GTNTX">GTN Textiles</a></td>
<td class="Number">46.35</td>
<td><span class="Number">42.10</span></td>
</tr>, <tr>
<td>4</td>
<td class="left"><a href="/company/AGUNIVERSAL">AG Universal</a></td>
<td class="Number">13.01</td>
<td><span class="Number">13.01</span></td>
</tr>, <tr>
<td>5</td>
<td class="left"><a href="/company/ASPIRAPATHLABDIAGNOSTICS">Aspira Pathlab&Diagn</a></td>
<td class="Number">21.60</td>
<td><span class="Number">21.60</span></td>
</tr>, <tr>
<td>6</td>
<td class="left"><a href="/company/SELLWINTRADERS">Sellwin Traders</a></td>
<td class="Number">19.22</td>
<td><span class="Number">19.22</span></td>
</tr>, <tr>
<td>7</td>
<td class="left"><a href="/company/SONIMEDICARE">Soni Medicare</a></td>
<td class="Number">7.85</td>
<td><span class="Number">7.85</span></td>
</tr>, <tr>
<td>8</td>
<td class="left"><a href="/company/SERAINVESTMENTSFINANCE">Sera Inve
stments&Fin</a></td>
<td class="Number">46.35</td>
<td><span class="Number">46.35</span></td>
</tr>, <tr>
<td>9</td>
<td class="left"><a href="/company/VIAZTYRES">Viaz Tyres</a></td>
<td class="Number">13.00</td>
<td><span class="Number">13.00</span></td>
</tr>, <tr>
<td>10</td>
<td class="left"><a href="/company/VIVANZABIOSCIENCES">Vivanza
Biosciences</a></td>
<td class="Number">5.55</td>
<td><span class="Number">5.55</span></td>
</tr>, <tr>
<td>11</td>
<td class="left"><a href="/company/TECILCHEM">TECIL Chem & Hydro</a></td>
<td class="Number">11.99</td>
<td><span class="Number">11.99</span></td>
</tr>, <tr>
<td>12</td>
<td class="left"><a href="/company/MILGREYFINANCE">Milgrey Fin.&Invest</a></td>
<td class="Number">8.81</td>
<td><span class="Number">8.81</span></td>
</tr>, <tr>
<td>13</td>
<td class="left"><a href="/company/ZODIACJRD">Zodiac-JRD-MK
J</a></td>
<td class="Number">16.05</td>
<td><span class="Number">16.05</span></td>
</tr>, <tr>
<td>14</td>
<td class="left"><a href="/company/KANUNGOFINANCIERS">Kanungo Financiers</a></td>
<td class="Number">5.35</td>
<td><span class="Number">5.35</span></td>
</tr>, <tr>
<td>15</td>
<td class="left"><a href="/company/CILNOVAPETRO">CIL Nova Petro</a></td>
<td class="Number">14.10</td>
<td><span class="Number">14.10</span></td>
</tr>, <tr>
<td>16</td>
<td class="left"><a href="/company/INTEGRATEDPRO">Integ.Pro</a></td>
<td class="Number">12.01</td>
<td><span class="Number">12.01</span></td>
</tr>, <tr>
<td>17</td>
<td class="left"><a href="/company/VOLTAIRELEASING">Voltaire Leasing</a></td>
<td class="Number">24.80</td>
<td><span class="Number">24.80</span></td>
</tr>, <tr>
<td>18</td>
<td class="left"><a href="/company/ARIHANTFOUNDATIONS">Arihant F
oundn. &Hsg</a></td>
<td class="Number">24.35</td>
<td><span class="Number">24.35</span></td>
</tr>, <tr>
<td>19</td>
<td class="left"><a href="/company/GLOBALOFFSHORESERVICES">Global Offshore Serv</a></td>
<td class="Number">7.75</td>
<td><span class="Number">7.75</span></td>
</tr>, <tr>
<td>20</td>
<td class="left"><a href="/company/ELANGOINDS">Elango Inds</a></td>
<td class="Number">17.40</td>
<td><span class="Number">17.40</span></td>
</tr>, <tr>
<td>21</td>
<td class="left"><a href="/company/AGUNIVERSAL">AG Universal</a></td>
<td class="Number">38.50</td>
<td><span class="Number">38.50</span></td>
</tr>, <tr>
<td>22</td>
<td class="left"><a href="/company/ASPIRAPATHLABDIAGNOSTICS">Aspira Pathlab&Diagn</a></td>
<td class="Number">8.33</td>
<td><span class="Number">8.33</span></td>
</tr>, <tr>
<td>23</td>
<td class="left"><a href="/company/SELLWINTRADERS">Sellwin Traders</a></td>
<td class="Number">11.50</td>
<td><span class="Number">11.50</span></td>
</tr>, <tr>
<td>24</td>
<td class="left"><a href="/company/SONIMEDICARE">Soni Medicare</a></td>
<td class="Number">20.70</td>
<td><span class="Number">20.70</span></td>
</tr>, <tr>
<td>25</td>
<td class="left"><a href="/company/SERAINVESTMENTSFINANCE">Sera Inve
stments&Fin</a></td>
<td class="Number">15.25</td>
<td><span class="Number">15.25</span></td>
</tr>, <tr>
<td>26</td>
<td class="left"><a href="/company/VIAZTYRES">Viaz Tyres</a></td>
<td class="Number">11.54</td>
<td><span class="Number">11.54</span></td>
</tr>]
```

```
In [67]: sr = soup.find_all("tr")
for i in sr[1:]:
    b=i.text[0:3].strip()
    dic["sno."].append(b)
print(dic)
```

```
{'sno.': ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19', '20', '21', '22', '23', '1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19', '20', '21', '22', '23'], 'Company': ['Rachana Infrastru', 'Tirupati Forge', 'Sera Investments&Fin', 'Viaz Tyres', 'Sellwin Traders', 'Soni Medicare', 'Sicagen India', 'Aspira Pathlab&Diagn', 'Patspin India', 'AG Universal', 'Arihant Foundn. &Hsg', 'Global Offshore Serv', 'GTN Textiles', 'GTN Inds', 'Vivanza Biosciences', 'TECIL Chem & Hydro', 'Milgrey Fin.&Invest', 'Zodiac-JRD-MKJ', 'Kanungo Financiers', 'CIL Nova Petro', 'Integ.Pro', 'Elango Inds', 'Voltaire Leasing'], 'Price Rs : ': ['125.70', '7.80', '13.01', '46.35', '13.00', '19.22', '24.35', '24.85', '14.10', '42.10', '39.65', '7.85', '19.50', '11.95', '8.33', '21.60', '16.29', '31.95', '5.35', '16.05', '8.81', '5.55', '11.99'], 'Day low Rs.': ['113.80', '7.75', '13.01', '45.00', '12.01', '17.40', '24.00', '24.80', '14.00', '42.00', '38.50', '7.75', '19.50', '11.50', '8.33', '20.70', '14.81', '31.50', '4.85', '15.25', '8.81', '5.55', '11.54']}
```

```
In [66]: import pandas as pd
df = pd.DataFrame(dic)
print(df)
```

	sno.	Company	Price Rs :	Day low Rs.
0	1	Rachana Infrastru	125.70	113.80
1	2	Tirupati Forge	7.80	7.75
2	3	Sera Investments&Fin	13.01	13.01
3	4	Viaz Tyres	46.35	45.00
4	5	Sellwin Traders	13.00	12.01
5	6	Soni Medicare	19.22	17.40
6	7	Sicagen India	24.35	24.00
7	8	Aspira Pathlab&Diagn	24.85	24.80
8	9	Patspin India	14.10	14.00
9	10	AG Universal	42.10	42.00
10	11	Arihant Foundn. &Hsg	39.65	38.50
11	12	Global Offshore Serv	7.85	7.75
12	13	GTN Textiles	19.50	19.50
13	14	GTN Inds	11.95	11.50
14	15	Vivanza Biosciences	8.33	8.33
15	16	TECIL Chem & Hydro	21.60	20.70
16	17	Milgrey Fin.&Invest	16.29	14.81
17	18	Zodiac-JRD-MKJ	31.95	31.50
18	19	Kanungo Financiers	5.35	4.85
19	20	CIL Nova Petro	16.05	15.25
20	21	Integ.Pro	8.81	8.81
21	22	Elango Inds	5.55	5.55
22	23	Voltaire Leasing	11.99	11.54

PB : 351


```
In [84]: import bs4
f = open("Certified used Mercedes-Benz for sale.html", encoding="utf8")
soup = bs4.BeautifulSoup(f, "html.parser")
print(soup.prettify())
```

```
<!DOCTYPE html>
<html class="ep-theme-spark" lang="en">
  <head>
    <!-- s: no | l: no -->
    <base __cpgenerated="1" href="https://www.cars.com/shopping/results/?s
tock_type=cpo&amp;makes%5B%5D=mercedes_benz&amp;models%5B%5D=&amp;list_p
rice_max=&amp;maximum_distance=20&amp;zip="/>
    <!-- CP_UI_TAG -->
    <script __cpp="1">
      window.__Cpn = window.__Cpn ? window.__Cpn : function () { this.perma
link = this.URI('https://stealthsurfer.online/__cp_i.php?s=Q05rb0oxUjA5Wm
tybzZ6LzBIOGt0bEJaZzJBNOsvUXVhRElWSTBoRXNFenVlUFFpa01YN25NY1Z6blNqb1lyVU
5KajgrSDVzZXNTTGY3OTc2cWRXMmZVeImIzS0M5RFp6U015QnIvT21VUFk9&r=aHR0cHM6Ly9
zdGVhbHRoc3VyZmVmYm9ubGlucyZS9zaG9wcGUyZy9yZXN1bHRZLz9zdG9ja190eXB1PWNwbyZ
tYWtlcyU1QiU1RD1tZXJjZWRLc19iZW56Jm1vZGVscyU1QiU1RD0mbGlzdF9wcmllZDV9tYXg
9Jm1heGltdW1fZGlzdGFuY2U9MjAmemlwPSZFx2Nwbz1hSFiwY0hNNkx5OTNkM2N1WTJGeWN
5NWpiMjA%3D&__cpo=''); this.modal = '<style __cpp="1"> #__cpsModal #__cp
sModalContent { background: url("<
UhEUgAAADIAAAAYCAAAAAAEPA4ixAAAAMjkLEQVRogV3awXLjSAwEUN7/tNy3Ion5CLyabE2Eo
...')';
```

```
In [85]: dic = {"car Name":[], "mileage":[], "Dealer Name":[], "Review":[], "Price":[]}
```

```
In [86]: cn = soup.find_all("h2",class_="title")
         for i in cn:
             b = i.text.strip()
             dic["car Name"].append(b)
```

```
In [87]: ml = soup.find_all("div", class_="mileage")
for i in ml[1:]:
    b = i.text.strip()
    dic["mileage"].append(b)
```

```
In [88]: dn = soup.find_all("div",class_="dealer-name")
for i in dn:
    b = i.text.strip()
    dic["Dealer Name"].append(b)
```

```
In [89]: re = soup.find_all("span",class_="sds-rating__link sds-button-link")
for i in re:
    b = i.text.strip("()reviews")
    dic["Review"].append(b)
```

```
In [90]: pr = soup.find_all("span", class_="primary-price")
         for i in pr:
             b = i.text.strip()
             dic["Price"].append(b)
```

```
In [91]: import pandas as pd
df = pd.DataFrame(dic)
df.head()
```

```
Out[91]:
```

	car Name	mileage	Dealer Name	Review	Price
0	2020 Mercedes-Benz CLA 250 Base 4MATIC	28,744 mi.	Mercedes-Benz of Lynnwood	130	\$35,995
1	2019 Mercedes-Benz AMG GT 53 Base	25,771 mi.	International Autos	117	\$79,995
2	2021 Mercedes-Benz AMG GLE 53 Base	22,374 mi.	John Sisson Motors	38	\$80,923
3	2022 Mercedes-Benz AMG CLA 45 Base 4MATIC	10,595 mi.	Mercedes-Benz of Santa Rosa	30	\$56,633
4	2021 Mercedes-Benz AMG GLE 53 Base	33,622 mi.	Mercedes-Benz of Rochester	152	\$76,995

```
In [ ]: # API KEY :669ccb11b84117c0c3ed6551cebb8d20
```