

```
#pip install beautifulsoup4
#pip install lxml -- if required
#pip install requests
#pip install selenium
#pip install webdriver-manager

# dataset -- google search --
# imdb top 250 movies
# amazon toys

from selenium import webdriver
from bs4 import BeautifulSoup

driver=webdriver.Chrome()
#driver.get('https://m.imdb.com/chart/top/?ref_=nv_mv_250')
driver.get("https://www.imdb.com/chart/top/")

content=driver.page_source
content

→ <html lang="en-US" xmlns:og="http://opengraphprotocol.org/schema/" xmlns:fb="http://www.facebook.com/2008/fbml" class=" scriptsOn">
<head><script async="" src="https://sb.scorecardresearch.com/beacon.js"></script><script async="" src="https://images-na.ssl-images-amazon.com/images/I/31bJewCvY-L.js"> crossorigin="anonymous"></script><meta charset="utf-8"><meta name="viewport" content="width=device-width"><script>if(typeof uet === 'function'){ uet('bb', 'LoadTitle\', {wb: 1}); }</script><title>IMDb Top 250 Movies</title><meta name="description" content="As rated by regular IMDb voters." data-id="main"><meta property="og:site_name" content="IMDb"><meta property="og:title" content="IMDb Top 250 Movies"><meta property="og:description" content="As rated by regular IMDb voters."><meta property="og:type" content="website"><meta property="og:image" content="https://m.media-amazon.com/images/G/01/imdb/images/social/imdb_logo.png"><meta property="og:image:height" content="1000"><meta property="og:image:width" content="1000"><meta content="en_US" property="og:locale"><meta content="es_ES" property="og:locale:alternate"><meta content="es_MX" property="og:locale:alternate"><meta content="fr_FR" property="og:locale:alternate"><meta content="fr_CA" property="og:locale:alternate"><meta content="it_IT" property="og:locale:alternate"><meta content="pt_BR" property="og:locale:alternate"><meta content="hi_IN" property="og:locale:alternate"><meta content="de_DE" property="og:locale:alternate"><meta property="twitter:site" content="@IMDb"><meta property="twitter:title" content="IMDb Top 250 Movies"><meta property="twitter:description" content="As rated by regular IMDb voters."><meta property="twitter:card" content="summary_large_image"><meta property="twitter:image" content="https://m.media-amazon.com/images/G/01/imdb/images/social/imdb_logo.png"><meta property="twitter:image:alt" content="As rated by regular IMDb voters."><meta property="imdb:pageType" content="chart"><meta property="imdb:subPageType" content="top250movie"><script>if(typeof uet === 'function'){ uet('be', 'LoadTitle\', {wb: 1}); }</script><script>if(typeof uex === 'function'){ uex('ld+', 'LoadTitle\', {wb: 1}); }</script><script type="application/ld+json">{"@type": "ItemList", "itemListElement": [{"@type": "ListItem", "item": {"@type": "Movie", "url": "https://www.imdb.com/title/tt0111161/", "name": "The Shawshank Redemption", "description": "A Maine banker convicted of the murder of his wife and her lover in 1947 gradually forms a friendship over a quarter century with a hardened convict, while maintaining his innocence and trying to remain hopeful through simple comp...", "image": "https://m.media-amazon.com/images/M/MV5BNDE30DcxYzMtY2YzZC00NmN1LWj1NDMtZDVizWM2MzTxZDYwXkEyXkFqcGdeQXVvNjAwNDUxODI@._V1_.jpg", "aggregateRating": {"@type": "AggregateRating", "bestRating": 10, "worstRating": 1, "ratingValue": 9.3, "ratingCount": 2921209}, "contentRating": "R", "genre": "Drama"}, {"@type": "ListItem", "item": {"@type": "Movie", "url": "https://www.imdb.com/title/tt0068646/", "name": "The Godfather", "description": "Don Vito Corleone, head of a mafia family, decides to hand over his empire to his youngest son, Michael. However, his decision unintentionally puts the lives of his loved ones in grave danger.", "image": "https://m.media-amazon.com/images/M/MV5BM2MyNjYxNmUtYTAwNi00MTYxLWjNWyTyzZL0Dy3ZTk30TF1XKEYXkFqcGdeQXVvNzkwmj05NzM@._V1_.jpg", "aggregateRating": {"@type": "AggregateRating", "bestRating": 10, "worstRating": 1, "ratingValue": 9.2, "ratingCount": 2035705}, "contentRating": "R", "genre": "Crime Drama", "duration": "PT2H55M"}, {"@type": "ListItem", "item": {"@type": "Movie", "url": "https://www.imdb.com/title/tt0468569/", "name": "The Dark Knight", "description": "When the menace known as the Joker wreaks havoc and chaos on the people of Gotham, Batman, Gordon and Harvey Dent work to stop the menace", "image": "https://m.media-amazon.com/images/M/MV5BMTMxNTMwODM0NF5B15BnBnXkFtZTcwODAyMTk2Mw@._V1_.jpg", "aggregateRating": {"@type": "AggregateRating", "bestRating": 10, "worstRating": 1, "ratingValue": 9, "ratingCount": 2902129}, "contentRating": "PG-13", "genre": "Action, Crime, Drama", "duration": "PT2H32M"}, {"@type": "ListItem", "item": {"@type": "Movie", "url": "https://www.imdb.com/title/tt0071562/", "name": "The Godfather Part II", "description": "The early life and career of Vito Corleone in 1920s New York City is portrayed, while his son, Michael, expands and tightens his grip on the family crime syndicate.", "image": "https://m.media-amazon.com/images/M/MV5BMwZTIt2J1NCO00WZlWIyMDctNDk2ZDQ2YjRjMNQ0XKEYXkFqcGdeQXVvNzkwmj05NzM@._V1_.jpg", "aggregateRating": {"@type": "AggregateRating", "bestRating": 10, "worstRating": 1, "ratingValue": 9, "ratingCount": 1377659}, "contentRating": "R", "genre": "Crime, Drama", "duration": "PT3H22M"}, {"@type": "ListItem", "item": {"@type": "Movie", "url": "https://www.imdb.com/title/tt0050083/", "name": "12 Angry Men", "description": "The jury in a New York City murder trial is frustrated by a single member whose skeptical caution forces them to more carefully consider the evidence before jumping to a hasty verdict.", "image": "https://m.media-amazon.com/images/M/MV5BMW4N2FjNzYtNTVkc00NzQ0LTg0MjAtYTj1MjFhNGUxZDFmXkEyXkFqcGdeQXVvNjc1NTYyMjg@._V1_.jpg", "aggregateRating": {"@type": "AggregateRating", "bestRating": 10, "worstRating": 1, "ratingValue": 9, "ratingCount": 876442}, "contentRating": "Approved", "genre": "Drama", "duration": "PT1H36M"}, {"@type": "ListItem", "item": {"@type": "Movie", "url": "https://www.imdb.com/title/tt0108052/", "name": "Schindler's List", "description": "In German-occupied Poland during World War II, industrialist Oskar Schindler gradually becomes concerned for his Jewish workforce after witnessing their persecution by the Nazis.", "image": "https://m.media-amazon.com/images/M/MV5BNDE40TMxMTctNmRhYy00NWE2Ltg3YzItYTk3M2UwOTU5Njg4XkEyXkFqcGdeQXVvNju00TO00TY@._V1_.jpg", "aggregateRating": {"@type": "AggregateRating", "bestRating": 10, "worstRating": 1, "ratingValue": 9, "ratingCount": 1465879}, "contentRating": "R", "genre": "Biography, Drama, History", "duration": "PT3H15M"}, {"@type": "ListItem", "item": {"@type": "Movie", "url": "https://www.imdb.com/title/tt0167260/", "name": "The Lord of the Rings: The Return of the King", "description": "Gandalf and Aragorn lead the World of Men against Sauron's army to draw his gaze from Frodo and Sam as they approach Mount Doom with the One Ring.", "image": "https://m.media-amazon.com/images/M/MV5BNzA5ZDN1ZWmt2NhsN00NDj1Ltk4NDItYTRmY2EwMWz1MTY3XkEvXkFqcGdeQXVvNzkwmj05NzM@._V1_.jpg", "aggregateRating": {"@type": "AggregateRating", "bestRating": 10, "worstRating": 1, "ratingValue": 9, "ratingCount": 1000000}, "contentRating": "PG-13", "genre": "Fantasy, Adventure, Drama", "duration": "PT10H30M"}]
```

```
soup=BeautifulSoup(driver.page_source,'html.parser')
```

```
soup.find_all("h3")
```

```
[<h3 class="ipc-title__text">IMDb Charts</h3>,<h3 class="ipc-title__text">1. The Shawshank Redemption</h3>,<h3 class="ipc-title__text">2. The Godfather</h3>,<h3 class="ipc-title__text">3. The Dark Knight</h3>,<h3 class="ipc-title__text">4. The Godfather Part II</h3>,<h3 class="ipc-title__text">5. 12 Angry Men</h3>,<h3 class="ipc-title__text">6. Schindler's List</h3>,<h3 class="ipc-title__text">7. The Lord of the Rings: The Return of the King</h3>,<h3 class="ipc-title__text">8. Pulp Fiction</h3>,<h3 class="ipc-title__text">9. The Lord of the Rings: The Fellowship of the Ring</h3>,<h3 class="ipc-title__text">10. The Good, the Bad and the Ugly</h3>,<h3 class="ipc-title__text">11. Forrest Gump</h3>,<h3 class="ipc-title__text">12. The Lord of the Rings: The Two Towers</h3>,<h3 class="ipc-title__text">13. Fight Club</h3>,<h3 class="ipc-title__text">14. Inception</h3>,<h3 class="ipc-title__text">15. Star Wars: Episode V - The Empire Strikes Back</h3>,<h3 class="ipc-title__text">16. The Matrix</h3>,<h3 class="ipc-title__text">17. Goodfellas</h3>,<h3 class="ipc-title__text">18. One Flew Over the Cuckoo's Nest</h3>,<h3 class="ipc-title__text">19. Interstellar</h3>,<h3 class="ipc-title__text">20. Se7en</h3>,<h3 class="ipc-title__text">21. It's a Wonderful Life</h3>,<h3 class="ipc-title__text">22. Seven Samurai</h3>,<h3 class="ipc-title__text">23. The Silence of the Lambs</h3>,<h3 class="ipc-title__text">24. Saving Private Ryan</h3>,<h3 class="ipc-title__text">25. City of God</h3>,<h3 class="ipc-title__text">26. Life Is Beautiful</h3>,<h3 class="ipc-title__text">27. The Green Mile</h3>,<h3 class="ipc-title__text">28. Terminator 2: Judgment Day</h3>,<h3 class="ipc-title__text">29. Star Wars: Episode IV - A New Hope</h3>,<h3 class="ipc-title__text">30. Back to the Future</h3>,<h3 class="ipc-title__text">31. Spirited Away</h3>,<h3 class="ipc-title__text">32. The Pianist</h3>,<h3 class="ipc-title__text">33. Parasite</h3>,<h3 class="ipc-title__text">34. Psycho</h3>,<h3 class="ipc-title__text">35. Gladiator</h3>,<h3 class="ipc-title__text">36. The Lion King</h3>,<h3 class="ipc-title__text">37. Spider-Man: Across the Spider-Verse</h3>,<h3 class="ipc-title__text">38. Dune: Part Two</h3>,<h3 class="ipc-title__text">39. The Departed</h3>,<h3 class="ipc-title__text">40. Léon: The Professional</h3>,<h3 class="ipc-title__text">41. American History X</h3>,<h3 class="ipc-title__text">42. Whiplash</h3>,<h3 class="ipc-title__text">43. The Prestige</h3>,<h3 class="ipc-title__text">44. Grave of the Fireflies</h3>,<h3 class="ipc-title__text">45. Harakiri</h3>,<h3 class="ipc-title__text">46. The Usual Suspects</h3>,<h3 class="ipc-title__text">47. Casablanca</h3>,<h3 class="ipc-title__text">48. The Intouchables</h3>,<h3 class="ipc-title__text">49. Cinema Paradiso</h3>,<h3 class="ipc-title__text">50. Modern Times</h3>,<h3 class="ipc-title__text">51. Rear Window</h3>,<h3 class="ipc-title__text">52. Once Upon a Time in the West</h3>,<h3 class="ipc-title__text">53. Alien</h3>,<h3 class="ipc-title__text">54. City Lights</h3>,<h3 class="ipc-title__text">55. Django Unchained</h3>,<h3 class="ipc-title__text">56. Apocalypse Now</h3>,<h3 class="ipc-title__text">57. Memento</h3>,
```

```
title_anchors=soup.find_all('a', attrs={'class':'ipc-title-link-wrapper'})  
title_anchors
```

```
[<a class="ipc-title-link-wrapper" href="/title/tt0111161/?ref_=chttp_t_1" tabindex="0"><h3 class="ipc-title__text">1. The Shawshank  
Redemption</h3></a>,<a class="ipc-title-link-wrapper" href="/title/tt0068646/?ref_=chttp_t_2" tabindex="0"><h3 class="ipc-title__text">2. The  
Godfather</h3></a>,<a class="ipc-title-link-wrapper" href="/title/tt0468569/?ref_=chttp_t_3" tabindex="0"><h3 class="ipc-title__text">3. The Dark  
Knight</h3></a>,<a class="ipc-title-link-wrapper" href="/title/tt0071562/?ref_=chttp_t_4" tabindex="0"><h3 class="ipc-title__text">4. The Godfather  
Part II</h3></a>,<a class="ipc-title-link-wrapper" href="/title/tt0050083/?ref_=chttp_t_5" tabindex="0"><h3 class="ipc-title__text">5. 12 Angry  
Men</h3></a>,<a class="ipc-title-link-wrapper" href="/title/tt0108052/?ref_=chttp_t_6" tabindex="0"><h3 class="ipc-title__text">6. Schindler's  
List</h3></a>,<a class="ipc-title-link-wrapper" href="/title/tt0167260/?ref_=chttp_t_7" tabindex="0"><h3 class="ipc-title__text">7. The Lord of  
the Rings: The Return of the King</h3></a>,<a class="ipc-title-link-wrapper" href="/title/tt0110912/?ref_=chttp_t_8" tabindex="0"><h3 class="ipc-title__text">8. Pulp  
Fiction</h3></a>,
```

```

<a class="ipc-title-link-wrapper" href="/title/tt0120737/?ref_=chttp_t_9" tabindex="0"><h3 class="ipc-title__text">9. The Lord of the Rings: The Fellowship of the Ring</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0060196/?ref_=chttp_t_10" tabindex="0"><h3 class="ipc-title__text">10. The Good, the Bad and the Ugly</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0109830/?ref_=chttp_t_11" tabindex="0"><h3 class="ipc-title__text">11. Forrest Gump</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0167261/?ref_=chttp_t_12" tabindex="0"><h3 class="ipc-title__text">12. The Lord of the Rings: The Two Towers</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0137523/?ref_=chttp_t_13" tabindex="0"><h3 class="ipc-title__text">13. Fight Club</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt1375666/?ref_=chttp_t_14" tabindex="0"><h3 class="ipc-title__text">14. Inception</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0080684/?ref_=chttp_t_15" tabindex="0"><h3 class="ipc-title__text">15. Star Wars: Episode V - The Empire Strikes Back</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0133093/?ref_=chttp_t_16" tabindex="0"><h3 class="ipc-title__text">16. The Matrix</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0099685/?ref_=chttp_t_17" tabindex="0"><h3 class="ipc-title__text">17. Goodfellas</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0073486/?ref_=chttp_t_18" tabindex="0"><h3 class="ipc-title__text">18. One Flew Over the Cuckoo's Nest</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0816692/?ref_=chttp_t_19" tabindex="0"><h3 class="ipc-title__text">19. Interstellar</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0114369/?ref_=chttp_t_20" tabindex="0"><h3 class="ipc-title__text">20. Se7en</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0038650/?ref_=chttp_t_21" tabindex="0"><h3 class="ipc-title__text">21. It's a Wonderful Life</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0047478/?ref_=chttp_t_22" tabindex="0"><h3 class="ipc-title__text">22. Seven Samurai</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0102926/?ref_=chttp_t_23" tabindex="0"><h3 class="ipc-title__text">23. The Silence of the Lambs</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0120815/?ref_=chttp_t_24" tabindex="0"><h3 class="ipc-title__text">24. Saving Private Ryan</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0317248/?ref_=chttp_t_25" tabindex="0"><h3 class="ipc-title__text">25. City of God</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0118799/?ref_=chttp_t_26" tabindex="0"><h3 class="ipc-title__text">26. Life Is Beautiful</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0120689/?ref_=chttp_t_27" tabindex="0"><h3 class="ipc-title__text">27. The Green Mile</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0103064/?ref_=chttp_t_28" tabindex="0"><h3 class="ipc-title__text">28. Terminator 2: Judgment Day</h3></a>,
<a class="ipc-title-link-wrapper" href="/title/tt0076759/?ref_=chttp_t_29" tabindex="0"><h3 class="ipc-title__text">29. Star Wars: Episode IV - A New Hope</h3></a>

```

```

title=[]
for a in title_anchors:
    title_h3=a.find('h3')
    t=title_h3.text
    #print(t)
    if t[0].isnumeric():
        t=t[t.find('.')+2:]
        #print(t)
        title.append(t)
title

```

→ ['The Shawshank Redemption',
 'The Godfather',
 'The Dark Knight',
 'The Godfather Part II',
 '12 Angry Men',
 "Schindler's List",
 'The Lord of the Rings: The Return of the King',
 'Pulp Fiction',
 'The Lord of the Rings: The Fellowship of the Ring',
 'The Good, the Bad and the Ugly',
 'Forrest Gump',
 'The Lord of the Rings: The Two Towers',
 'Fight Club',
 'Inception',
 'Star Wars: Episode V - The Empire Strikes Back',
 'The Matrix',
 'Goodfellas',
 "One Flew Over the Cuckoo's Nest",
 'Interstellar',
 'Se7en',
 "It's a Wonderful Life",
 'Seven Samurai',
 'The Silence of the Lambs',
 'Saving Private Ryan',
 'City of God',
 'Life Is Beautiful',
 'The Green Mile',
 'Terminator 2: Judgment Day',
 'Star Wars: Episode IV - A New Hope',

```
'Back to the Future',
'Spirited Away',
'The Pianist',
'Parasite',
'Psycho',
'Gladiator',
'The Lion King',
'Spider-Man: Across the Spider-Verse',
'Dune: Part Two',
'The Departed',
'Léon: The Professional',
'American History X',
'Whiplash',
'The Prestige',
'Grave of the Fireflies',
'Harakiri',
'The Usual Suspects',
'Casablanca',
'The Intouchables',
'Cinema Paradiso',
'Modern Times',
'Rear Window',
'Once Upon a Time in the West',
'Alien',
'City Lights',
'Django Unchained',
'Apocalypse Now',
'Memento',
```

```
year_div=soup.find_all('div',
                       attrs={'class':'sc-b189961a-7 btCcOY cli-title-metadata'})
```

```
year_div
```

```
[<div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1994</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 22m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">R</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1972</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 55m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">R</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2008</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 32m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">PG-13</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1974</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">3h 22m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">R</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1957</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1h 36m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">Approved</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1993</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">3h 15m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">R</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2003</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">3h 21m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">PG-13</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1994</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 34m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">R</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2001</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 58m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">PG-13</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1966</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 58m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">R</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1994</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 22m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">PG-13</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2002</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 59m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">PG-13</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1999</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 19m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">R</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2010</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 28m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">PG-13</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1980</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 4m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">PG</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1999</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 16m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">R</span></div>,
 <div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1990</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 25m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">
```

```
item">R</span></div>,
<div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">1975</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 13m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">R</span></div>,
<div class="sc-b189961a-7 btCcOY cli-title-metadata"><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2014</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">2h 49m</span><span class="sc-b189961a-8 hCbzGp cli-title-metadata-item">PG-13</span></div>,
...
...
...
year=[]
duration=[]
for d in year_div:
    span=d.find_all('span')
    year.append(span[0].text)
    duration.append(span[1].text)

year
→ ['1994',
 '1972',
 '2008',
 '1974',
 '1957',
 '1993',
 '2003',
 '1994',
 '2001',
 '1966',
 '1994',
 '2002',
 '1999',
 '2010',
 '1980',
 '1999',
 '1990',
 '1975',
 '2014',
 '1995',
 '1946',
 '1954',
 '1991',
 '1998',
 '2002',
 '1997',
 '1999',
 '1991',
 '1977',
 '1985',
 '2001',
 '2002',
 '2019',
 '1960',
 '2000',
 '1994',
 '2023',
 '2024',
 '2006',
 '1994',
 '1998',
 '2014',
 '2006',
 '1988',
 '1962',
 '1995',
 '1942',
 '2011',
 '1988',
 '1936',
 '1954',
 '1968',
 '1979',
 '1931',
 '2012',
 '1979',
 '2000',
 '2008',
 '2008']

import pandas as pd
df=pd.DataFrame({'Movie_Title':title,'Year':year,'Duration':duration})
df.head()
```

	Movie Title	Year	Duration
0	The Shawshank Redemption	1994	2h 22m
1	The Godfather	1972	2h 55m
2	The Dark Knight	2008	2h 32m
3	The Godfather Part II	1974	3h 22m
4	12 Angry Men	1957	1h 36m

```
df.to_csv("C:\\\\Users\\\\dell\\\\Documents\\\\ParinitaJain\\\\Python9-11\\\\imdb_scraped_data.csv")
```

```
driver=webdriver.Chrome()
driver.get('https://www.amazon.in/s?k=amazon+toys&i=toys&rh=n%3A1350380031%2Cp_n_deal_type%3A26921226031&dc&ds=v1%3AHkOCakvpCLC9XT4Ck9PFTym%
```

```
driver.page_source
```

```
<html lang="en-in" class=" a-js a-audio a-video a-canvas a-drag-drop a-geolocation a-history a-webworker a-autofocus a-input-placeholder a-textarea-placeholder a-local-storage a-gradients a-transform3d a-touch-scrolling a-text-stroke a-box-shadow a-border-radius a-border-image a-opacity a-transform a-transition null" data-19ax5a9jf="dingo" data-oui-build-date="3.24.5-2024-07-24"><!-- sp:feature:head-start --><head><script async="" src="https://c.amazon-adsystem.com/bao-csm/forensics/a9-tq-forensics-incremental_min.js" crossorigin="anonymous"></script><script async="" src="https://images-eu.ssl-images-amazon.com/images/I/31bJewCvY-L.js" crossorigin="anonymous"></script><script>var aPageStart = (new Date()).getTime();</script><meta charset="utf-8"><n!-- sp:end-feature:head-start --><n!-- sp:feature:csm:head-open-part1 --><n><script type="text/javascript">var ue_t0=ue_t0||+new Date();</script><n!-- sp:end-feature:csm:head-open-part1 --><n!-- sp:feature:cs-optimization --><n><meta http-equiv="x-dns-prefetch-control" content="on"><n><link rel="dns-prefetch" href="https://images-eu.ssl-images-amazon.com" crossorigin=""><n><link rel="dns-prefetch" href="https://m.media-amazon.com" crossorigin=""><n><link rel="dns-prefetch" href="https://completion.amazon.com" crossorigin=""><n><link rel="dns-prefetch" href="https://completion.amazon.com" crossorigin=""><n><link rel="dns-prefetch" href="https://completion.amazon.com" crossorigin=""><n><script type="text/javascript"><n>window.ue_ihb = (window.ue_ihb || window.ueinit || 0) + 1;<n>if (window.ue_ihb == 1) {<n>var ue_csm = window,<n>ue_hob = +new Date();<n>(function(d){var e=d.ue=d.ue||{},f=Date.now||function(){return+new Date();}e.d=function(b){return f}-(b?d.ue_t0)});e.stub=function(b,a){if(!b[a])var c=[];b[a]=function()<n>{c.push([c.slice.call(arguments),e.d(),d.ue_id])};b[a].replay=function(b){for(var a;a=c.shift();)b(a[0],a[1],a[2]);b[a].isStub=1};e.exec=function(b,a){return function(){try{return b.apply(this,arguments)}catch(c){ue.LogError(c,{attribution:a||"undefined",logLevel:"WARN"})}}}(ue_csm);<n>var ue_err_chan = '\jserr-rw';<n>(function(d,e){function h(f,b){if(!!(a.ec>a.mxe)&&f){a.ter.push(f);b=b||{};var c=f.logLevel||b.logLevel;c&&1==k&&c!=n&&c!=p||a.ec++&&c&c!=k||a.ecf++;b.pageURL=""+(e.location.href:"");b.logLevel=c;b.attribution=f.attribution||b.attribution;a.erl.push({ex:f,info:b})}}function l(a,b,c,e,g){d.LogError({m:a,f:b,l:c,"":e,err:g,fromOnError:1,args:arguments},g?{attribution:g.attribution,logLevel:g.logLevel}:void 0);return!1}var k="FATAL",m="ERROR",p="DOWNGRADED",a={ec:0,ecf:0,\npec:0,ts:0,erl:[],ter:[],buffer:[],mxe:50,startTimer:function(){a.ts++;setInterval(function(){<n>{d.ue&&a.pec<&&d.ue(x("at"))&&a.pec=a.ec,1E4)};1.skipTrace=1;h.skipTrace=1;h.isStub=1;d.ue.LogError=h.d.ue_err=a;e.onerror=1})(ue_csm,window);<n>var ue_id = '\1YDBVHH28Y8ZEFK07NZD\',<n>ue_url = '/rd/uedata\',<n>ue_navtiming = 1,<n>ue_mid = '\A21TJRUUN4KGV\',<n>ue_sid = '\259-5223595-6731468\',<n>ue_sn = '\www.amazon.in\',<n>ue_furl = '\fls-eu.amazon.in\',<n>ue_surl = '\https://unagi-eu.amazon.com/1/events/com.amazon.csm.nexusclient.prod\',<n>ue_int = 0,<n>ue_fcsn = 1,<n>ue_urt = 3,<n>ue_rpl_ns = '\cel-rpl1\',<n>ue_ddq = 1,<n>ue_fpf = '\//fls-eu.amazon.in/1/batch/1/OP/A21TJRUUN4KGV:259-5223595-6731468:1YDBVHH28Y8ZEFK07NZD#uedata=s:\',<n>ue_sbuimp = 1,<n>ue_ibft = 0,<n>ue_sswmts = 0,<n>ue_jsmtf = 0,<n>ue_fnt = 0,<n>ue_lpsi = 6000,<n>ue_no_counters = 1,<n>ue_lob = '\1\',<n>ue_sjslob = 0,<n>ue_sw1 = 1;<n>var ue_viz=function(){(function(b,f,d){function g(){return(!(p in d)||0<d[p]&&!(q in d)||0<d[q])}function h(c){if(b.ue.viz.length<=&&r){var a=c.type;c=c.originalEvent;/^focus/.test(a)&&c&&c.toElement||c.fromElement||c.relatedTarget)|||a(g)?f[s]||("blur"==a||"focusout"==a?t:u):t,b.ue.viz.push(a:+(+new Date-b.ue.t0)),a=u&&(b.ue.is1&&x("at"),r=1)}}for(var r=0,x=b.uxex,a,k,l,s,v=[[],"webkit","o","ms","moz"],e=0,m=1,u="visible",t="hidden",p="innerWidth",q="innerHeight",w=20,n=0;n<v.length&&e;n++)if(a=\nv[n],k=(a?"+H":")+"+idden",e="boolean"==typeof f[k])=a+"visibilitychange",s=(a?"+V":"v")+"isibilityState";h());e&&f.addEventListener(1,h,0);m=g()?1:0;d.addEventListener("resize",function(){var a=g()?1:0;m!=a&&(m=a,h({}),{passive:!0});b.ue&&e&&(b.ue.pageViz={event:1,propHid:k})})(ue_csm,ue_csm.document,ue_csm.window);<n>(function(d,h,N){function H(a){return a&&a.replace&&a.replace(/\\"+\\"/\\"+$\+/g,"")function u(a){return"undefined"==typeof a}function B(a,b){for(var c in b)b[v](c)&&(a[c]=b[c])}function I(a){try{var b=N.cookie.match(RegExp("(^| )"+a+"=([^;]+)"));}catch(e){}}function O(k,b,c){var q=x||{};type;if("device"!=c||!2==q&&1!=q){k&&(d.ue_id=a.id=a.rid=k,w&w.replace(/((.*?:(?2){2})\(\w+/,function(a,b){return b+k}),D&&(e("id",D,k,D=0)),b&&(w&&(w=w.replace(/((.*?:(\w+)-)/,function(a,\nc){return c+b}),d.ue_sid=b),c&&a.tag("page-source":+c),d.ue_fp=w)}function P(){var a={};return function(b){b&&(a[b]=1);b=[];for(var c in a)a[v](c)&&b.push(c);return b}}function y(d,b,c,q){q=q||+new E;var g,m;if(b|u(c)){if(d)for(m in g=b?e("t",b)|e("t",b,{}):a.t,g[d]=q,c[v](m)&&e(m,b,c[m]);return q}}function e(d,b,c){var e=b&&!a.id?a.sc[b]:a.e||(e=a.sc[b]={});"id"==d&&c&&(Q=1);return e[d]=c||e[d]}function R(d,b,c,e,g){var m=b;c,g="on"+c;var n=m[b];"function"==typeof m?d&&(a.h[d]=m=function(){})&&b[c]=\nfunction(a){g?(e(a),m(a)):((m(a),e(a));b[c]&&(b[c].isUeh=1))}function S(k,b,c,q){function p(b,c){var d=[b],f=0,g={},m,h;c?(d.push("m=1"),g[c]=1):g.a.sc;for(h in g)if(g[v](h)){var q=e("wb",h),p=e("t",h)||{},n=e("t0",h)||a.t0,1;if(c||2==q){q=q?f++:"";}d.push("sc"+q+"="+h);for(l in p)u(p[l])||null==p[l]||d.push(l+q+"="+p[l]-n);d.push("t"+q+"="+p[k]);if(e("ctb",h)|e("wb",h))m=1}}!38&m&d.push("ctb=1");return d.join("&")}function m(b,c,f,e,g){if(b){var k=d.ue_err;d.ue_url&&e!=g&&b&&0<b.length&&(e=\nnew Image,a.iel.push(e),e.src=b,a.count&&a.count("postbackImageSize",b.length));}w?(g=h.encodeURIComponent)&&&(e=new Image,b=""+d.ue_fp+g(b)+"":(+new E-d.ue_t0),a.iel.push(e),e.src=b):a.log&&(a.log(b,"uedata",{n:1}),a.ielf.push(b));k&&!k.ts&&k.startTimer();a.b&&(k=a.b,a.b=","+m(k,c,f,1))}function A(b){var c=x?
```

```
soup=BeautifulSoup(driver.page_source,'html.parser')
```

```
product_span_tag = soup.find_all('span', attrs={'class': 'a-size-base-plus a-color-base a-text-normal'})  
product_name=[p.text for p in product_span_tag]  
product_name  
  
→ ['Amazon Brand - Jam & Honey | 4 Level Ramp Race Gift for Kids | 2 Cars',  
 'Amazon Brand - Jam & Honey Suitcase - Makeup Set Multicolor',  
 'Amazon Brand - Jam & Honey Penguin, Plush/Soft Toy for Boys, Girls and Kids, Super-Soft, Safe, Great Birthday Gift (Black and White, 17 cm)',  
 'Amazon Brand - Solimo Penguin Slide with Basketball for Indoor & Outdoor Use | Safe & Fun | Ideal for Boys & Girls | Toy for kids | Easy Assembly | Suitable for age 24M-4Y',  
 'Toyzone Baby Panda Magic Car-50940 | Ride-on Baby Car| Kids Push Car| Swing Car | Comfortable Seats & Durable | Elegant Design | Ride on Toy Car for Kids|Twister Ride on|Magic Toy Car|Kids Ride On|Push Rider|Steering Music & Lights | For Kids age 2+',  
 'Amazon Brand - Jam & Honey Map of India Puzzle - Foam material with Flags for State Capitals | Educational Toy | for Boys & Girls | for Ages 5 yrs & above | 24 Pieces (Colors Assorted)',  
 'Little's Lighthouse Stacking Drums | Activity Toy for Babies I Multicolor I Infant & Preschool Toys I Develops Motor & Reasoning Skills',  
 'Amazon Brand - Jam & Honey Hugging Elephant | Stuffed Animal Soft Toy of Plush Hugging Pillow for Kids Boys and Girls | Non-Toxic & Safe | BIS-Approved (Grey)',  
 'Amazon Brand - Jam & Honey Polyester Baby Bowling Ball/Soft Plush Toy with Rattle Sound for Boys, Girls and Kids, Super-Soft, Safe, Great Birthday Gift (Multicoloured)',  
 'ZUDO TOYS Shape Sorting Cube With 18 Shape And Different Color - Kids Activity Toys - Learning And Educational Toys With Multicolour - (Bis Approved), 19 Pcs',  
 'ZUDO TOYS Learning Colour And Pattern Matching Montessori Slide Puzzle Game For Kids, Multicolor',  
 'Storio Duck Slide Track Set Toys for Kids - Ducklings Stair Climbing Toys for Kids, Escalator Toy with Lights and Music with 3 Ducks',  
 'Amazon Brand - Jam & Honey Plastic Suitcase - Kitchen Set for Kids',  
 'Amazon Brand - Jam & Honey Indoor Outdoor Ping Pong Table Tennis Trainer | Sports Toy for Teenagers & Kids | Birthday Gift''',  
 'Amazon Brand - Jam & Honey Teddy Bear, Cute, Plush/Soft Toy for Boys, Girls and Kids, Super-Soft, Safe, Great Birthday Gift (Multicoloured, 28 cm)',  
 'Amazon Brand - Jam & Honey Kid - Friction Toy|Dumper Truck,Yellow',  
 'Amazon Brand - Jam & Honey Dolphin, Plush/Soft Toy for Boys, Girls and Kids, Super-Soft, Safe, Great Birthday Gift (Blue, 18 cm)',  
 'Amazon Brand - Jam & Honey Penguin, Plush/Soft Toy for Boys, Girls and Kids, Super-Soft, Safe, Great Birthday Gift (Pink and White, 17 cm)',  
 'Amazon Brand - Jam & Honey Capital Alphabet Tray with Picture (with Knobs)',  
 'Amazon Brand - Jam & Honey Drum Keyboard | Musical Toy with Flashing Lights | Animal Sounds | Fun Toy for Kids and Parties',  
 'Amazon Brand - Jam & Honey Kid - Friction Toy|Concrete Truck,Yellow',  
 'Amazon Brand - Jam & Honey Plastic Learning & Activity Ball I Learning Activity Toy I Multicolour I Infant and Preschool Toys | Good Toy for Parties and Return Gifts, Handball, Standard',  
 'Amazon Brand - Jam & Honey High Contrast Flash Cards for Babies & Toddlers | Newborn Visual Stimulation | Cognitive, Visual & Language Development Tools | Educational Toy | BIS Approved (Set of 30)',  
 'Amazon Brand - Jam & Honey Glittery Eyes Cat, Plush/Soft/Cute Toy for Boys, Girls and Kids, Great Birthday Gift (17 cm)',  
 'Amazon Brand - Jam & Honey Duck, Plush/Soft Toy for Boys, Girls and Kids, Super-Soft, Safe, Great Birthday Gift (Yellow, 18 cm)',  
 'Amazon Brand - Jam & Honey 4 in 1 Adventure - Puzzles for Kids | 35-Piece Puzzles | Educational Toy for Cognitive Development | Vibrant Colors',  
 'Amazon Brand - Jam & Honey Dinosaur, Cute, Plush/Soft Toy, Suitable for Boys, Girls and Kids, Super-Soft, Safe, 34 cm (Green & Yellow)',  
 'Amazon Brand - Jam & Honey Teddy Bear, Cute, Soft Toy (33 Cm, Pink), Great Birthday Gift',  
 'Babiique Unicorn Teddy Bear Plush Soft Toy Cute Kids Birthday Animal Baby Boys/Girls (25 Cm, Yellow)',  
 'Toyzone Baby Panda Magic Car-50940 | Ride-on Baby Car| Kids Push Car| Swing Car | Comfortable Seats & Durable | Elegant Design | Ride on Toy Car for Kids|Twister Ride on|Magic Toy Car|Kids Ride On|Push Rider|Steering Music & Lights | For Kids age 2+',  
 'Amazon Brand - Jam & Honey Reversible Carrot Strawberry Bunny Plush Toy | Zipper | Cute Stuffed Animal | Rabbit Soft Toy |Age 3 Years and Above',  
 'Amazon Brand - Jam & Honey Foldable Toy Storage Box for Kids',  
 'Amazon Brand - Jam & Honey Penguin, Plush/Soft Toy for Boys, Girls and Kids, Super-Soft, Safe, Great Birthday Gift (Pink and White, 17 cm)',  
 'Amazon Brand - Jam & Honey Husky Dog, Plush/Soft Toy for Boys, Girls and Kids, Super-Soft, Safe, Great Birthday Gift (Grey and White, 17 cm)']
```

```
price_span_tag = soup.find_all('span', attrs={'class': 'a-price-whole'})  
product_price = [p.text.replace(',', '') for p in price_span_tag]  
product_price
```

```
→ ['359',  
 '389',  
 '169',  
 '1999',  
 '1198',  
 '299',  
 '174',  
 '529',  
 '125',  
 '240',  
 '280',  
 '429',  
 '379',  
 '525',  
 '280',  
 '359',  
 '309',  
 '160',  
 '169',  
 '269',  
 '525',
```

```
'249',
'279',
'399',
'189',
'199',
'249',
'225',
'335',
'169',
'1198',
'399',
'949',
'169',
'199']
```

```
rating_span_tag=soup.find_all('span', attrs={'class':'a-icon-alt'})
product_rating = [p.text[:3] for p in rating_span_tag[4:-1]]
product_rating
```

```
→ ['4.4',
'4.1',
'4.0',
'4.0',
'4.1',
'4.1',
'4.1',
'4.1',
'4.2',
'4.2',
'3.5',
'4.0',
'4.0',
'4.0',
'4.2',
'4.5',
'3.9',
'4.1',
'4.4',
'3.7',
'4.1',
'3.9',
'4.1',
'4.8',
'4.4',
'4.0',
'4.0',
'4.4',
'4.4',
'4.3',
'4.1',
'4.0',
'4.4',
'5.0',
'4.4',
'4.1']
```

```
print(len(product_name),len(product_price),len(product_rating))
product_name=product_name[:33]
product_price=product_price[:33]
```

```
→ 34 34 33
```

```
df=pd.DataFrame({'title':product_name,
                 'price':product_price,
                 'rating':product_rating})
df.to_csv("C:\\\\Users\\\\dell\\\\Documents\\\\ParinitaJain\\\\ParinitaJain9_11\\\\amazon_toys_scraped_data.csv",index=False)
```

Start coding or generate with AI.

