

ANALYST: NIRAJAN RIJAL ; Linkedin: <https://www.linkedin.com/in/nirajan-rijal-8b1766274/>

github: <https://github.com/NirajanRijal1>

```
In [1]: from autoscraper import AutoScraper
```

```
In [2]: url="https://hdghar.com/"
```

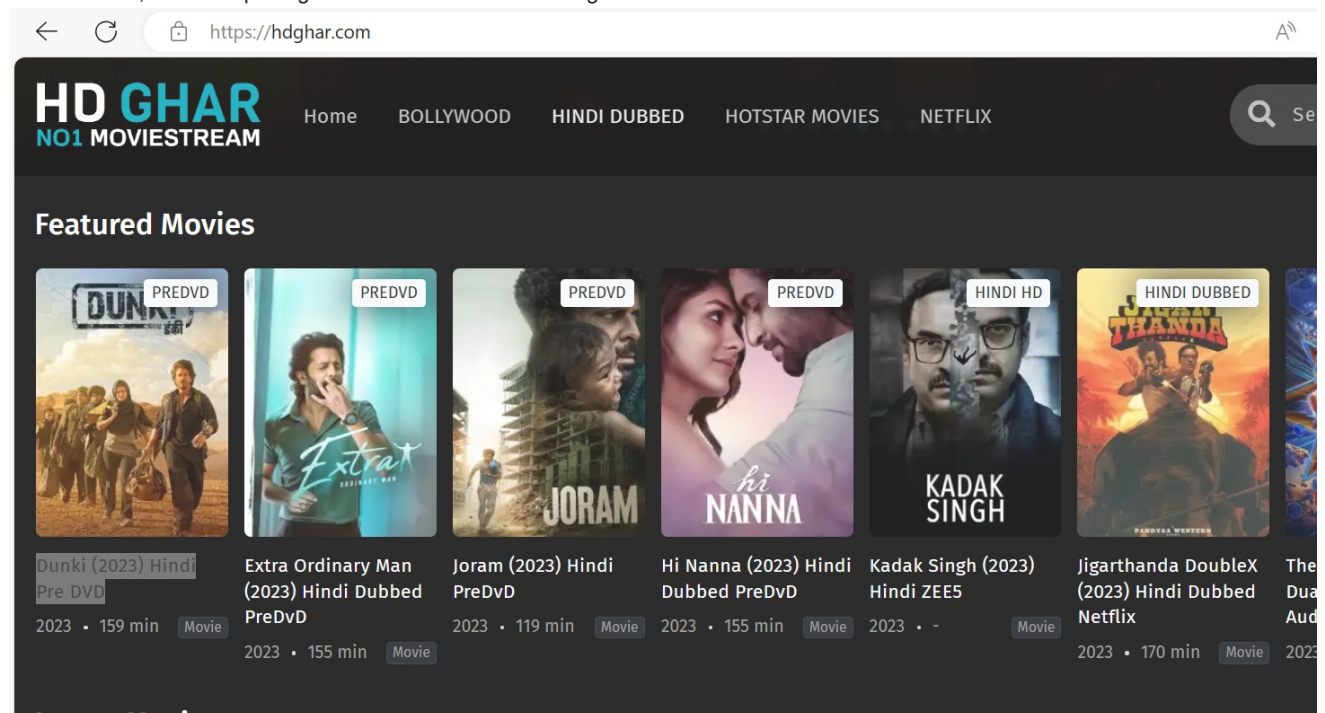
```
In [3]: #adding candidate - we can add one or more candidate
wanted_list=["Dunki (2023) Hindi Pre DVD","159 min"] # currently we are adding movie name and length of movie
```

```
In [4]: scraper=AutoScraper() #Build object for AutoScraper class
result=scraper.build(url,wanted_list)
```

```
In [5]: print(result)
```

```
['Dunki (2023) Hindi Pre DVD', 'Extra Ordinary Man (2023) Hindi Dubbed PreDvD', 'Joram (2023) Hindi PreDvD', 'Hi Nanna (2023) Hindi Dubbed PreDvD', 'Kadak Singh (2023) Hindi ZEE5', 'Jigarthanda DoubleX (2023) Hindi Dubbed Netflix', 'The Marvels (2023) Dual [Hindi Line Audio]', 'Parinda Paar Geyaa (2023) Punjabi Pre DVD', 'Yu Yu Hakusho (2023) Hindi Season 1 Complete Netflix', 'Art of Fighting 1 (2019) Hindi Dubbed', 'Minus 31-The Nagpur Files (2023) Hindi HD', 'Under Pressure: The U.S. Womens World Cup Team (2023) Hindi Season 1 Complete Netflix', 'Los Farad (2023) Hindi Season 1 Complete', 'Japan (2023) Hindi Dubbed Netflix', 'Baby Ruby (2023) Hindi Dubbed', 'Captain Marvel (2019) Hindi Dubbed', 'Avatar (2009) Hindi Dubbed', 'Ratsasan (2020) Hindi Dubbed', 'King Richard (2021) Hindi Dubbed', 'Any How Mitti Pao (2023) Punjabi HD', 'Sheroes (2023) Hindi Dubbed', 'Zakir Khan Mannpasand (2023) Stand Up Comedy', 'Adbhut (2023) Hindi Season 1 Complete', 'My Girl (1991) Hindi Dubbed', 'Planet of the Apes (2001) Hindi Dubbed', 'Madhagaja (2021) Hindi Dubbed', 'The Spy Who Dumped Me (2018) Hindi Dubbed', 'SAS: Red Notice (2021) Hindi Dubbed', 'Room 203 (2022) Hindi Dubbed', 'HIT: The First Case (2020) Hindi Dubbed', 'Iron Man 3 (2013) Hindi Dubbed', 'Leave the World Behind (2023) Hindi Dubbed Netflix', 'The Moon (2023) Hindi Dubbed', 'Mast Mein Rehne Ka (2023) Hindi HD', 'Silver and the Book of Dreams (2023) Hindi Dubbed', 'Detective Knight: Redemption (2022) Hindi Dubbed', '159 min', '155 min', '119 min', '-', '170 min', '105 min', '134 min', '154 min', '93 min', '124 min', '162 min', '145 min', '91 min', '94 min', '103 min', '120 min', '132 min', '117 min', '104 min', '126 min', '130 min', '140 min', '129 min', '127 min', '92 min']
```

For reference, we are importing above information and storing in result from below website:



Now we should have some python knowledge

```
In [15]: import pandas as pd
import re

df = pd.DataFrame(result, columns=['title'])
print(df)
```

	title	
0	Dunki (2023) Hindi Pre DVD	
1	Extra Ordinary Man (2023) Hindi Dubbed PreDvD	
2	Joram (2023) Hindi PreDvD	
3	Hi Nanna (2023) Hindi Dubbed PreDvD	
4	Kadak Singh (2023) Hindi ZEE5	
5	Jigarthanda DoubleX (2023) Hindi Dubbed Netflix	
6	The Marvels (2023) Dual [Hindi Line Audio]	
7	Parinda Paar Geyaa (2023) Punjabi Pre DVD	
8	Yu Yu Hakusho (2023) Hindi Season 1 Complete N...	
9	Art of Fighting 1 (2019) Hindi Dubbed	
10	Minus 31-The Nagpur Files (2023) Hindi HD	
11	Under Pressure: The U.S. Womens World Cup Team...	
12	Los Farad (2023) Hindi Season 1 Complete	
13	Japan (2023) Hindi Dubbed Netflix	
14	Baby Ruby (2023) Hindi Dubbed	
15	Captain Marvel (2019) Hindi Dubbed	
16	Avatar (2009) Hindi Dubbed	
17	Ratsasan (2020) Hindi Dubbed	
18	King Richard (2021) Hindi Dubbed	
19	Any How Mitti Pao (2023) Punjabi HD	
20	Sheroes (2023) Hindi Dubbed	
21	Zakir Khan Mannpasand (2023) Stand Up Comedy	
22	Adbhut (2023) Hindi Season 1 Complete	
23	My Girl (1991) Hindi Dubbed	
24	Planet of the Apes (2001) Hindi Dubbed	
25	Madhagaja (2021) Hindi Dubbed	
26	The Spy Who Dumped Me (2018) Hindi Dubbed	
27	SAS: Red Notice (2021) Hindi Dubbed	
28	Room 203 (2022) Hindi Dubbed	
29	HIT: The First Case (2020) Hindi Dubbed	
30	Iron Man 3 (2013) Hindi Dubbed	
31	Leave the World Behind (2023) Hindi Dubbed Net...	
32	The Moon (2023) Hindi Dubbed	
33	Mast Mein Rehne Ka (2023) Hindi HD	
34	Silver and the Book of Dreams (2023) Hindi Dubbed	
35	Detective Knight: Redemption (2022) Hindi Dubbed	
36		159 min
37		155 min
38		119 min
39		-
40		170 min
41		105 min
42		134 min
43		154 min
44		93 min
45		124 min
46		162 min
47		145 min
48		91 min
49		94 min
50		103 min
51		120 min
52		132 min
53		117 min
54		104 min
55		126 min
56		130 min
57		140 min
58		129 min
59		127 min
60		92 min

```
In [9]: df.head()
```

```
Out[9]:
```

	title
0	Dunki (2023) Hindi Pre DVD
1	Extra Ordinary Man (2023) Hindi Dubbed PreDvD
2	Joram (2023) Hindi PreDvD
3	Hi Nanna (2023) Hindi Dubbed PreDvD
4	Kadak Singh (2023) Hindi ZEE5

```
In [16]: print(df)
```

	title	
0	Dunki (2023) Hindi Pre DVD	
1	Extra Ordinary Man (2023) Hindi Dubbed PreDvD	
2	Joram (2023) Hindi PreDvD	
3	Hi Nanna (2023) Hindi Dubbed PreDvD	
4	Kadak Singh (2023) Hindi ZEE5	
5	Jigarthanda DoubleX (2023) Hindi Dubbed Netflix	
6	The Marvels (2023) Dual [Hindi Line Audio]	
7	Parinda Paar Geyaa (2023) Punjabi Pre DVD	
8	Yu Yu Hakusho (2023) Hindi Season 1 Complete N...	
9	Art of Fighting 1 (2019) Hindi Dubbed	
10	Minus 31-The Nagpur Files (2023) Hindi HD	
11	Under Pressure: The U.S. Womens World Cup Team...	
12	Los Farad (2023) Hindi Season 1 Complete	
13	Japan (2023) Hindi Dubbed Netflix	
14	Baby Ruby (2023) Hindi Dubbed	
15	Captain Marvel (2019) Hindi Dubbed	
16	Avatar (2009) Hindi Dubbed	
17	Ratsasan (2020) Hindi Dubbed	
18	King Richard (2021) Hindi Dubbed	
19	Any How Mitti Pao (2023) Punjabi HD	
20	Sheroes (2023) Hindi Dubbed	
21	Zakir Khan Mannpasand (2023) Stand Up Comedy	
22	Adbhut (2023) Hindi Season 1 Complete	
23	My Girl (1991) Hindi Dubbed	
24	Planet of the Apes (2001) Hindi Dubbed	
25	Madhagaja (2021) Hindi Dubbed	
26	The Spy Who Dumped Me (2018) Hindi Dubbed	
27	SAS: Red Notice (2021) Hindi Dubbed	
28	Room 203 (2022) Hindi Dubbed	
29	HIT: The First Case (2020) Hindi Dubbed	
30	Iron Man 3 (2013) Hindi Dubbed	
31	Leave the World Behind (2023) Hindi Dubbed Net...	
32	The Moon (2023) Hindi Dubbed	
33	Mast Mein Rehne Ka (2023) Hindi HD	
34	Silver and the Book of Dreams (2023) Hindi Dubbed	
35	Detective Knight: Redemption (2022) Hindi Dubbed	
36		159 min
37		155 min
38		119 min
39		-
40		170 min
41		105 min
42		134 min
43		154 min
44		93 min
45		124 min
46		162 min
47		145 min
48		91 min
49		94 min
50		103 min
51		120 min
52		132 min
53		117 min
54		104 min
55		126 min
56		130 min
57		140 min
58		129 min
59		127 min
60		92 min

```
In [17]: df.to_csv(r"C:\Users\USER\Desktop\saved\saved.csv", index=False) #if want to save as csv in local computer
```

```
In [24]: #I edit some portion in excel, now importing it in Pandas
```

```
import numpy as np
import pandas as pd
```

```
In [31]: df_updated=pd.read_csv(r"C:\Users\USER\Desktop\saved\saved.csv")
```

```
In [32]: df_updated.head()
```

```
Out[32]:
```

	title	length
0	Dunki (2023) Hindi Pre DVD	159 min
1	Extra Ordinary Man (2023) Hindi Dubbed PreDvD	155 min
2	Joram (2023) Hindi PreDvD	119 min
3	Hi Nanna (2023) Hindi Dubbed PreDvD	NaN
4	Kadak Singh (2023) Hindi ZEE5	170 min

```
In [33]: df_updated.isna()
```

Out[33]:

	title	length
0	False	False
1	False	False
2	False	False
3	False	True
4	False	False
5	False	False
6	False	False
7	False	False
8	False	False
9	False	False
10	False	False
11	False	False
12	False	False
13	False	False
14	False	False
15	False	False
16	False	False
17	False	False
18	False	False
19	False	False
20	False	False
21	False	False
22	False	False
23	False	False
24	False	False
25	False	True
26	False	True
27	False	True
28	False	True
29	False	True
30	False	True
31	False	True
32	False	True
33	False	True
34	False	True
35	False	True

In [35]: df\_updated.info()

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 36 entries, 0 to 35  
Data columns (total 2 columns):  
# Column Non-Null Count Dtype  
--- ---  
0 title 36 non-null object  
1 length 24 non-null object  
dtypes: object(2)  
memory usage: 708.0+ bytes

In [36]: df\_updated.describe()

Out[36]:

	title	length
count	36	24
unique	36	24
top	Dunki (2023) Hindi Pre DVD	159 min
freq	1	1