In [1]:
```
pip install requests
```

```
Requirement already satisfied: requests in c:\users\admin\anaconda3\lib\site-packages (2.26.0)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\admin\anaconda3\lib\site-packages (from requests) (2021.1
0.8)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\admin\anaconda3\lib\site-packages (from requests) (1.2
6.7)
Requirement already satisfied: idna<4,>=2.5 in c:\users\admin\anaconda3\lib\site-packages (from requests) (3.2)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\admin\anaconda3\lib\site-packages (from requests)
(2.0.4)
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```
pip install html5lib
```

```
Requirement already satisfied: html5lib in c:\users\admin\anaconda3\lib\site-packages (1.1)
Requirement already satisfied: six>=1.9 in c:\users\admin\anaconda3\lib\site-packages (from html5lib) (1.16.0)
Requirement already satisfied: webencodings in c:\users\admin\anaconda3\lib\site-packages (from html5lib) (0.5.1)
Note: you may need to restart the kernel to use updated packages.
```

In [3]:
```
pip install selenium
```

```
Requirement already satisfied: sniffio in c:\users\admin\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.
2.0)
Requirement already satisfied: idna in c:\users\admin\anaconda3\lib\site-packages (from trio~=0.17->selenium) (3.2)
Requirement already satisfied: async-generator>=1.9 in c:\users\admin\anaconda3\lib\site-packages (from trio~=0.17->s
elenium) (1.10)
Requirement already satisfied: sortedcontainers in c:\users\admin\anaconda3\lib\site-packages (from trio~=0.17->selen
ium) (2.4.0)
Requirement already satisfied: pycparser in c:\users\admin\anaconda3\lib\site-packages (from cffi>=1.14->trio~=0.17->
selenium) (2.20)
Requirement already satisfied: wsproto>=0.14 in c:\users\admin\anaconda3\lib\site-packages (from trio-websocket~=0.9-
>selenium) (1.1.0)
Requirement already satisfied: pyOpenSSL>=0.14 in c:\users\admin\anaconda3\lib\site-packages (from urllib3[secure,soc
ks]~=1.26->selenium) (21.0.0)
Requirement already satisfied: cryptography>=1.3.4 in c:\users\admin\anaconda3\lib\site-packages (from urllib3[secur
e,socks]~=1.26->selenium) (3.4.8)
Requirement already satisfied: certifi in c:\users\admin\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.2
6->selenium) (2021.10.8)
Requirement already satisfied: PySocks!=1.5.7,<2.0,>=1.5.6 in c:\users\admin\anaconda3\lib\site-packages (from urllib
3[secure,socks]~=1.26->selenium) (1.7.1)
Requirement already satisfied: six>=1.5.2 in c:\users\admin\anaconda3\lib\site-packages (from pyOpenSSL>=0.14->urllib
```

In [4]:
```
pip install lxml
```

```
Requirement already satisfied: lxml in c:\users\admin\anaconda3\lib\site-packages (4.6.3)
Note: you may need to restart the kernel to use updated packages.
```

In [1]:
```
import requests
import pandas as pd
from bs4 import BeautifulSoup
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
import numpy as np
import re
from datetime import datetime
```

In [47]:
```
url='https://www.themoviedb.org/movie'
```

In [3]:
```
url = 'https://www.themoviedb.org'

header = {'User-Agent': 'Mozilla/5.0 (Windows NT 6.3; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/99.0.4844

response = requests.get(url+'/movie', headers = header)
```

In [4]:
```
response
```

Out[4]:
```
<Response [200]>
```

In [5]:
```
content=response.text
```

In [6]:
```
soup = BeautifulSoup(content,'lxml')
```

In [7]:
```
card_style = soup.find('div',class_='card style_1')
movie_name= card_style.find('h2').text
movie_rating= soup.find('div',class_='user_score_chart')['data-percent']
movie_link='https://www.themoviedb.org'+card_style.find('a')['href']
url2=requests.get(movie_link,headers=header)
movie_url=BeautifulSoup(url2.text,'lxml')
movie_release=movie_url.find('span',class_='release').text
movie_genres=movie_url.find('span',class_='genres').text.strip().replace('\xa0','')
movie_director=movie_url.find('li',class_='profile').text
movie_genres
```

Out[7]:
```
'Crime,Mystery,Thriller'
```

In [8]:
```
card_style
```

Out[8]:
```
<div class="card style_1">
<div class="image">
<div class="wrapper">
<a class="image" href="/movie/414906" title="The Batman">
<img alt="" class="poster" loading="lazy" src="/t/p/w220_and_h330_face/74xTEgt7R36Fpooo50r9T25onhq.jpg" srcset="/t/p/
w220_and_h330_face/74xTEgt7R36Fpooo50r9T25onhq.jpg 1x, /t/p/w440_and_h660_face/74xTEgt7R36Fpooo50r9T25onhq.jpg 2x"/>
</a>
</div>
<div class="options" data-id="414906" data-media-type="movie" data-object-id="57d1b0c49251410cae0003fd">
<a class="no_click" href="#"><div class="glyphicons_v2 circle-more white"></div></a>
</div>
</div>
<div class="content">
<div class="consensus tight">
<div class="outer_ring">
<div class="user_score_chart 57d1b0c49251410cae0003fd" data-bar-color="#21d07a" data-percent="79.0" data-track-color
="#204529">
<div class="percent">
<span class="icon icon-r79"></span>
```

In [9]:
```
page_url='https://www.themoviedb.org/movie?page='
```

In [10]:
```
lst_url=[]
for value in range(1,201):
    lst_url.append(page_url+str(value))
```

In [11]:
```
for val in lst_url:
    print(val)
```

```
https://www.themoviedb.org/movie?page=1 (https://www.themoviedb.org/movie?page=1)
https://www.themoviedb.org/movie?page=2 (https://www.themoviedb.org/movie?page=2)
https://www.themoviedb.org/movie?page=3 (https://www.themoviedb.org/movie?page=3)
https://www.themoviedb.org/movie?page=4 (https://www.themoviedb.org/movie?page=4)
https://www.themoviedb.org/movie?page=5 (https://www.themoviedb.org/movie?page=5)
https://www.themoviedb.org/movie?page=6 (https://www.themoviedb.org/movie?page=6)
https://www.themoviedb.org/movie?page=7 (https://www.themoviedb.org/movie?page=7)
https://www.themoviedb.org/movie?page=8 (https://www.themoviedb.org/movie?page=8)
https://www.themoviedb.org/movie?page=9 (https://www.themoviedb.org/movie?page=9)
https://www.themoviedb.org/movie?page=10 (https://www.themoviedb.org/movie?page=10)
https://www.themoviedb.org/movie?page=11 (https://www.themoviedb.org/movie?page=11)
https://www.themoviedb.org/movie?page=12 (https://www.themoviedb.org/movie?page=12)
https://www.themoviedb.org/movie?page=13 (https://www.themoviedb.org/movie?page=13)
https://www.themoviedb.org/movie?page=14 (https://www.themoviedb.org/movie?page=14)
https://www.themoviedb.org/movie?page=15 (https://www.themoviedb.org/movie?page=15)
https://www.themoviedb.org/movie?page=16 (https://www.themoviedb.org/movie?page=16)
https://www.themoviedb.org/movie?page=17 (https://www.themoviedb.org/movie?page=17)
https://www.themoviedb.org/movie?page=18 (https://www.themoviedb.org/movie?page=18)
https://www.themoviedb.org/movie?page=19 (https://www.themoviedb.org/movie?page=19)
```

In [14]:
```
movie_list=[]
def get_all_movie_list():
    count=0
    for link in lst_url:
        count+=1
        print(count)
        response=requests.get(link)
        html_text=response.text
        movies_soup=BeautifulSoup(html_text,'lxml')
        movie=soup.find_all('div',class_='card style_1')
        for item in movie:
            movie_percent=item.find('div',class_='user_score_chart')['data-percent']
            movie_name=item.find('h2').text
            movie_link='https://www.themoviedb.org'+item.find('a')['href']
            url2=requests.get(movie_link,headers=header)
            mo_url=BeautifulSoup(url2.text,'lxml')
            mo_url
            movie_genres=mo_url.find('span',class_='genres').text.strip().replace('\xa0',' ')
            movie_release=mo_url.find('span',class_='release').text.strip()
            movie_runtime=mo_url.find('span',class_='runtime').text.strip()
            movie_director=mo_url.find('li',class_='profile').a.text
            my_movie = {
                'movie_name' : movie_name,
                'movie_percent' : movie_percent,
                'movie_genres' : movie_genres,
                'movie_release' : movie_release,
                'movie_runtime' : movie_runtime,
                'movie_director' : movie_director,
                'movie_link' : movie_link
                }
            movie_list.append(my_movie)
get_all_movie_list()
```

```
9
10
11
12
13
14
15
16
```

```
17
18
19
20
21
22
23
24
25
26
27
27
28
```

In [ ]:
```
df= pd.DataFrame(movie_list)
```

In [ ]:
```
df
```

In [ ]:
```
df.to_csv('movie_data.csv')
```

In [2]:
```
df= pd.read_csv('movie_data.csv')
```

In [3]:
```
df
```

Out[3]:

	Unnamed: 0	movie_name	movie_percent	movie_genres	movie_release	movie_runtime	movie_director	movie_lin
0	0	The Batman	79.0	Crime, Mystery, Thriller	03/04/2022 (IN)	2h 56m	Matt Reeves	https://www.themoviedb.org/movie/41490
1	1	The Outfit	72.0	Drama, Thriller, Crime	03/18/2022 (US)	1h 45m	Graham Moore	https://www.themoviedb.org/movie/79987
2	2	Spider-Man: No Way Home	81.0	Action, Adventure, Science Fiction	12/17/2021 (IN)	2h 28m	Steve Ditko	https://www.themoviedb.org/movie/63464
3	3	Turning Red	75.0	Animation, Family, Comedy, Fantasy	03/10/2022 (CZ)	1h 40m	Domee Shi	https://www.themoviedb.org/movie/50894
				Action, Science				

In [4]:
```
data=df.copy()
```

In [5]:
```
data
```

Out[5]:

	Unnamed: 0	movie_name	movie_percent	movie_genres	movie_release	movie_runtime	movie_director	movie_link
0	0	The Batman	79.0	Crime, Mystery, Thriller	03/04/2022 (IN)	2h 56m	Matt Reeves	https://www.themoviedb.org/movie/414906
1	1	The Outfit	72.0	Drama, Thriller, Crime	03/18/2022 (US)	1h 45m	Graham Moore	https://www.themoviedb.org/movie/799876
2	2	Spider-Man: No Way Home	81.0	Action, Adventure, Science Fiction	12/17/2021 (IN)	2h 28m	Steve Ditko	https://www.themoviedb.org/movie/634649
3	3	Turning Red	75.0	Animation, Family, Comedy, Fantasy	03/10/2022 (CZ)	1h 40m	Domee Shi	https://www.themoviedb.org/movie/508947
4	4	Sonic the Hedgehog 2	77.0	Action, Science Fiction, Comedy, Family	04/08/2022 (US)	2h 2m	Josh Miller	https://www.themoviedb.org/movie/675353
...	...	...	...	...	...	...	...	...
3995	3995	Gold	64.0	Thriller, Action	01/13/2022 (AU)	1h 37m	Anthony Hayes	https://www.themoviedb.org/movie/760926
3996	3996	Hotel Transylvania: Transformania	71.0	Animation, Family, Fantasy, Comedy, Adventure	02/25/2022 (US)	1h 27m	Genndy Tartakovsky	https://www.themoviedb.org/movie/585083
3997	3997	Restless	60.0	Action, Thriller, Crime	02/25/2022 (FR)	1h 35m	Régis Blondeau	https://www.themoviedb.org/movie/928381
3998	3998	Beautiful Sisters: Flesh Slave	52.0	Crime, Horror	01/18/1986 (JP)	1h 10m	Katsuhiko Fujii	https://www.themoviedb.org/movie/340553
3999	3999	The In Between	71.0	Romance, Science Fiction, Drama	02/11/2022 (US)	1h 56m	Arie Posin	https://www.themoviedb.org/movie/818750

4000 rows × 8 columns

In [10]:
```
data.drop(['Unnamed: 0'],axis=1,inplace=True)
```

In [11]:
```
data
```

Out[11]:

	movie_name	movie_percent	movie_genres	movie_release	movie_runtime	movie_director	movie_link
0	The Batman	79.0	Crime, Mystery, Thriller	03/04/2022 (IN)	2h 56m	Matt Reeves	https://www.themoviedb.org/movie/414906
1	The Outfit	72.0	Drama, Thriller, Crime	03/18/2022 (US)	1h 45m	Graham Moore	https://www.themoviedb.org/movie/799876
2	Spider-Man: No Way Home	81.0	Action, Adventure, Science Fiction	12/17/2021 (IN)	2h 28m	Steve Ditko	https://www.themoviedb.org/movie/634649
3	Turning Red	75.0	Animation, Family, Comedy, Fantasy	03/10/2022 (CZ)	1h 40m	Domee Shi	https://www.themoviedb.org/movie/508947
4	Sonic the Hedgehog 2	77.0	Action, Science Fiction, Comedy, Family	04/08/2022 (US)	2h 2m	Josh Miller	https://www.themoviedb.org/movie/675353
...	...	...	...	...	...	...	...
3995	Gold	64.0	Thriller, Action	01/13/2022 (AU)	1h 37m	Anthony Hayes	https://www.themoviedb.org/movie/760926
	Hotel		Animation, Family				

In [12]:
```
data.columns = ['Name', 'Rating', 'Genres', 'Date','Runtime', 'Director', 'link']
data
```

Out[12]:

	Name	Rating	Genres	Date	Runtime	Director	link
0	The Batman	79.0	Crime, Mystery, Thriller	03/04/2022 (IN)	2h 56m	Matt Reeves	https://www.themoviedb.org/movie/414906
1	The Outfit	72.0	Drama, Thriller, Crime	03/18/2022 (US)	1h 45m	Graham Moore	https://www.themoviedb.org/movie/799876
2	Spider-Man: No Way Home	81.0	Action, Adventure, Science Fiction	12/17/2021 (IN)	2h 28m	Steve Ditko	https://www.themoviedb.org/movie/634649
3	Turning Red	75.0	Animation, Family, Comedy, Fantasy	03/10/2022 (CZ)	1h 40m	Domee Shi	https://www.themoviedb.org/movie/508947
4	Sonic the Hedgehog 2	77.0	Action, Science Fiction, Comedy, Family	04/08/2022 (US)	2h 2m	Josh Miller	https://www.themoviedb.org/movie/675353
...	...	...	...	...	...	...	...
3995	Gold	64.0	Thriller, Action	01/13/2022 (AU)	1h 37m	Anthony Hayes	https://www.themoviedb.org/movie/760926
3996	Hotel Transylvania: Transformania	71.0	Animation, Family, Fantasy, Comedy, Adventure	02/25/2022 (US)	1h 27m	Genndy Tartakovsky	https://www.themoviedb.org/movie/585083
3997	Restless	60.0	Action, Thriller, Crime	02/25/2022 (FR)	1h 35m	Régis Blondeau	https://www.themoviedb.org/movie/928381
3998	Beautiful Sisters: Flesh Slave	52.0	Crime, Horror	01/18/1986 (JP)	1h 10m	Katsuhiko Fujii	https://www.themoviedb.org/movie/340553
3999	The In Between	71.0	Romance, Science Fiction, Drama	02/11/2022 (US)	1h 56m	Arie Posin	https://www.themoviedb.org/movie/818750

4000 rows × 7 columns

In [13]:
```
data['Date']=data['Date'].str[0:10]
data
```

Out[13]:

	Name	Rating	Genres	Date	Runtime	Director	link
0	The Batman	79.0	Crime, Mystery, Thriller	03/04/2022	2h 56m	Matt Reeves	https://www.themoviedb.org/movie/414906
1	The Outfit	72.0	Drama, Thriller, Crime	03/18/2022	1h 45m	Graham Moore	https://www.themoviedb.org/movie/799876
2	Spider-Man: No Way Home	81.0	Action, Adventure, Science Fiction	12/17/2021	2h 28m	Steve Ditko	https://www.themoviedb.org/movie/634649
3	Turning Red	75.0	Animation, Family, Comedy, Fantasy	03/10/2022	1h 40m	Domee Shi	https://www.themoviedb.org/movie/508947
4	Sonic the Hedgehog 2	77.0	Action, Science Fiction, Comedy, Family	04/08/2022	2h 2m	Josh Miller	https://www.themoviedb.org/movie/675353
...	...	...	...	...	...	...	...
3995	Gold	64.0	Thriller, Action	01/13/2022	1h 37m	Anthony Hayes	https://www.themoviedb.org/movie/760926
3996	Hotel Transylvania: Transformania	71.0	Animation, Family, Fantasy, Comedy, Adventure	02/25/2022	1h 27m	Genndy Tartakovsky	https://www.themoviedb.org/movie/585083

In [14]:
```
data.isnull().sum()
```

Out[14]:
```
Name        0
Rating      0
Genres      0
Date        0
Runtime     0
Director    0
link        0
dtype: int64
```

In [15]:
```
data.max()
```

Out[15]:
```
Name                    Yaksha: Ruthless Operations
Rating                                         81.0
Genres                                Thriller, Action
Date                                        12/22/2021
Runtime                                         2h 5m
Director                                   Steve Ditko
link        https://www.themoviedb.org/movie/928381 (https://www.themoviedb.org/movie/928381)
dtype: object
```

In [16]:
```
data.min()
```

Out[16]:
```
Name                           All the Old Knives
Rating                                       52.0
Genres                                     Action
Date                                     01/13/2022
Runtime                                      1h 10m
Director                                  Adam Berg
link        https://www.themoviedb.org/movie/294793 (https://www.themoviedb.org/movie/294793)
dtype: object
```

calculating average number of rating

In [17]:
```
data.groupby('Genres')['Rating'].agg(['count','mean'])
```

Out[17]:

Genres	count	mean
Action	200	62.0
Action, Adventure, Comedy, Science Fiction	200	70.0
Action, Adventure, Science Fiction	400	73.0
Action, Science Fiction, Comedy, Family	200	77.0
Action, Thriller	400	61.0
Action, Thriller, Crime	200	60.0
Adventure, Animation, Comedy, Family, Fantasy	200	68.0
Animation, Comedy, Family, Fantasy	200	77.0
Animation, Family, Comedy, Fantasy	200	75.0
Animation, Family, Fantasy, Comedy, Adventure	200	71.0
Crime, Horror	200	52.0
Crime, Mystery, Thriller	200	79.0
Drama, Thriller, Crime	200	72.0
Horror, Thriller	200	67.0
Romance, Science Fiction, Drama	200	71.0
Science Fiction, Action	200	60.0
Thriller, Action	400	62.0

In [18]:
```
data.groupby(['Name'])['Rating','Genres','Date','Runtime','Director'].min()
```

```
C:\Users\Admin\AppData\Local\Temp/ipykernel_3436/1553414436.py:1: FutureWarning: Indexing with multiple keys (implicitl
y converted to a tuple of keys) will be deprecated, use a list instead.
  data.groupby(['Name'])['Rating','Genres','Date','Runtime','Director'].min()
```

Out[18]:

Name	Rating	Genres	Date	Runtime	Director
All the Old Knives	60.0	Thriller, Action	04/08/2022	1h 42m	Olen Steinhauer
Beautiful Sisters: Flesh Slave	52.0	Crime, Horror	01/18/1986	1h 10m	Katsuhiko Fujii
Black Crab	62.0	Action, Thriller	03/18/2022	1h 49m	Adam Berg
Blacklight	60.0	Action, Thriller	02/11/2022	1h 44m	Mark Williams
Encanto	77.0	Animation, Comedy, Family, Fantasy	11/24/2021	1h 42m	Charise Castro Smith
Gold	64.0	Thriller, Action	01/13/2022	1h 37m	Anthony Hayes
Hotel Transylvania: Transformania	71.0	Animation, Family, Fantasy, Comedy, Adventure	02/25/2022	1h 27m	Genndy Tartakovsky
Moonfall	65.0	Action, Adventure, Science Fiction	02/04/2022	2h 10m	Roland Emmerich
No Exit	67.0	Horror, Thriller	02/25/2022	1h 36m	Damien Power
Pil's Adventures	68.0	Adventure, Animation, Comedy, Family, Fantasy	08/11/2021	1h 29m	Julien Fournet
Restless	60.0	Action, Thriller, Crime	02/25/2022	1h 35m	Régis Blondeau
Sonic the Hedgehog 2	77.0	Action, Science Fiction, Comedy, Family	04/08/2022	2h 2m	Josh Miller
Spider-Man: No Way Home	81.0	Action, Adventure, Science Fiction	12/17/2021	2h 28m	Stan Lee
The Adam Project	70.0	Action, Adventure, Comedy, Science Fiction	03/11/2022	1h 46m	Shawn Levy
The Batman	79.0	Crime, Mystery, Thriller	03/04/2022	2h 56m	Matt Reeves
The In Between	71.0	Romance, Science Fiction, Drama	02/11/2022	1h 56m	Arie Posin
The Outfit	72.0	Drama, Thriller, Crime	03/18/2022	1h 45m	Graham Moore
Turning Red	75.0	Animation, Family, Comedy, Fantasy	03/10/2022	1h 40m	Domee Shi
War of the Worlds: Annihilation	60.0	Science Fiction, Action	12/22/2021	1h 53m	Maximilian Elfeldt
Yaksha: Ruthless Operations	62.0	Action	04/08/2022	2h 5m	Na Hyun

In [19]:
```
data.groupby(['Name'])['Rating','Genres','Date','Runtime','Director'].max()
```

```
C:\Users\Admin\AppData\Local\Temp/ipykernel_3436/1984316264.py:1: FutureWarning: Indexing with multiple keys (implicitl
y converted to a tuple of keys) will be deprecated, use a list instead.
  data.groupby(['Name'])['Rating','Genres','Date','Runtime','Director'].max()
```

Out[19]:

Name	Rating	Genres	Date	Runtime	Director
All the Old Knives	60.0	Thriller, Action	04/08/2022	1h 42m	Olen Steinhauer
Beautiful Sisters: Flesh Slave	52.0	Crime, Horror	01/18/1986	1h 10m	Katsuhiko Fujii
Black Crab	62.0	Action, Thriller	03/18/2022	1h 49m	Adam Berg
Blacklight	60.0	Action, Thriller	02/11/2022	1h 44m	Mark Williams
Encanto	77.0	Animation, Comedy, Family, Fantasy	11/24/2021	1h 42m	Jared Bush
Gold	64.0	Thriller, Action	01/13/2022	1h 37m	Anthony Hayes
Hotel Transylvania: Transformania	71.0	Animation, Family, Fantasy, Comedy, Adventure	02/25/2022	1h 27m	Genndy Tartakovsky
Moonfall	65.0	Action, Adventure, Science Fiction	02/04/2022	2h 10m	Roland Emmerich
No Exit	67.0	Horror, Thriller	02/25/2022	1h 36m	Damien Power
Pil's Adventures	68.0	Adventure, Animation, Comedy, Family, Fantasy	08/11/2021	1h 29m	Julien Fournet
Restless	60.0	Action, Thriller, Crime	02/25/2022	1h 35m	Régis Blondeau
Sonic the Hedgehog 2	77.0	Action, Science Fiction, Comedy, Family	04/08/2022	2h 2m	Patrick Casey
Spider-Man: No Way Home	81.0	Action, Adventure, Science Fiction	12/17/2021	2h 28m	Steve Ditko
The Adam Project	70.0	Action, Adventure, Comedy, Science Fiction	03/11/2022	1h 46m	Shawn Levy
The Batman	79.0	Crime, Mystery, Thriller	03/04/2022	2h 56m	Matt Reeves
The In Between	71.0	Romance, Science Fiction, Drama	02/11/2022	1h 56m	Arie Posin
The Outfit	72.0	Drama, Thriller, Crime	03/18/2022	1h 45m	Graham Moore
Turning Red	75.0	Animation, Family, Comedy, Fantasy	03/10/2022	1h 40m	Domee Shi
War of the Worlds: Annihilation	60.0	Science Fiction, Action	12/22/2021	1h 53m	Maximilian Elfeldt
Yaksha: Ruthless Operations	62.0	Action	04/08/2022	2h 5m	Na Hyun

In [20]:
```
data.describe()
```

Out[20]:

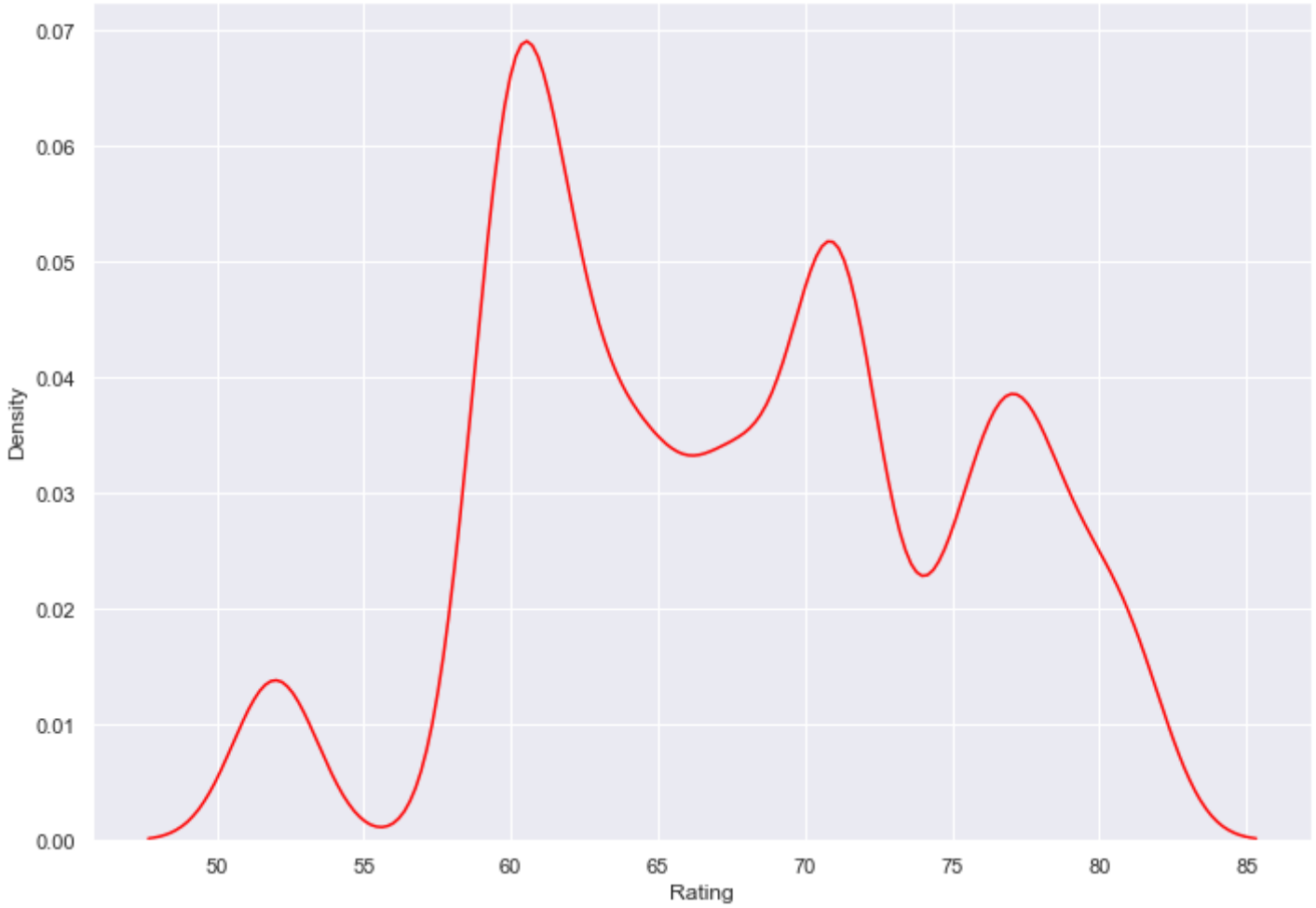
	Rating
count	4000.000000
mean	67.650000
std	7.572439
min	52.000000
25%	61.500000
50%	67.500000
75%	72.750000
max	81.000000

In [214]:
```
sns.distplot(data['Rating'],hist=False,color='red')
```

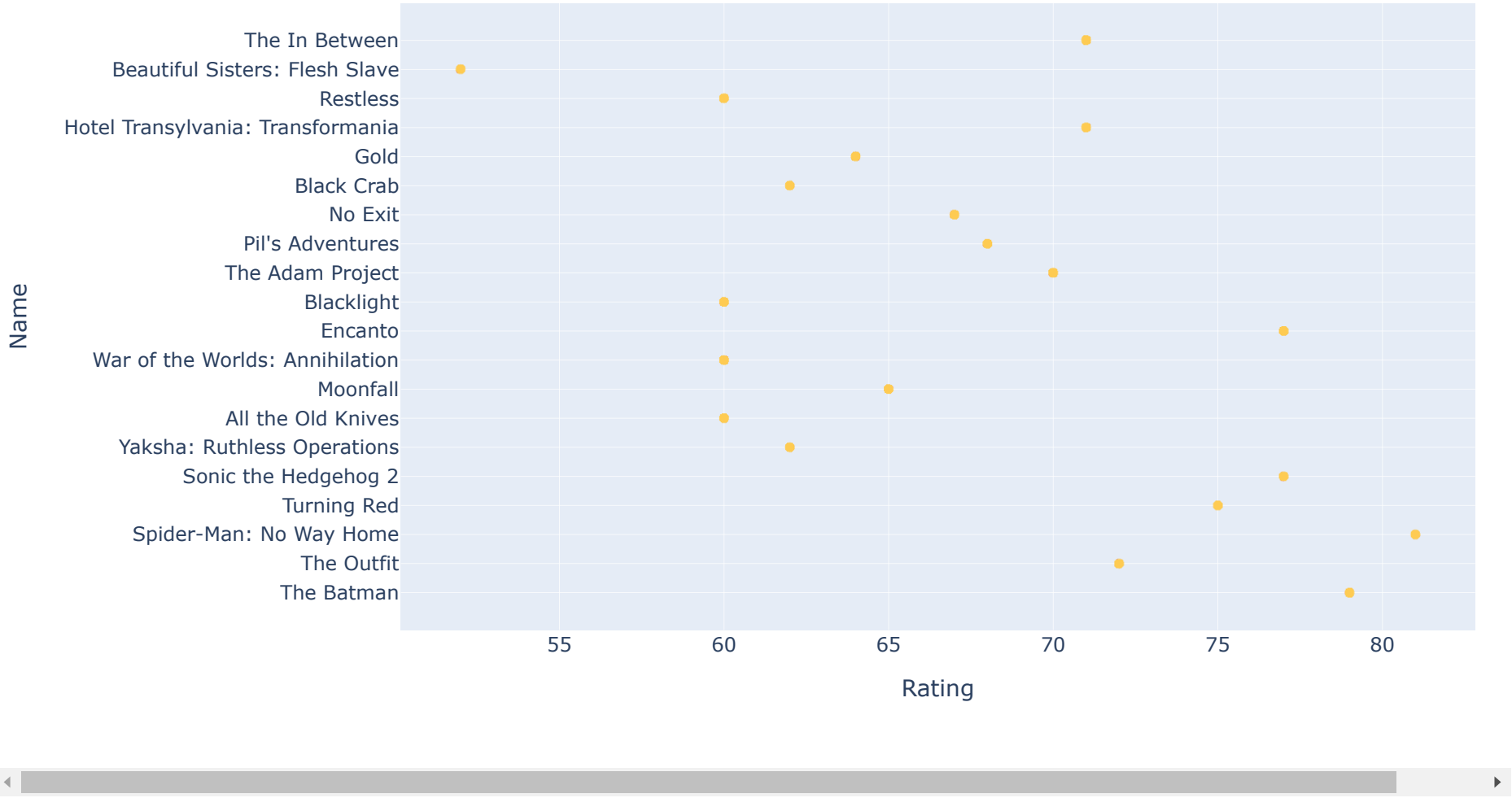
```
C:\Users\Admin\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning:

`distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `kdeplot` (an axes-level function for kernel density plots).
```

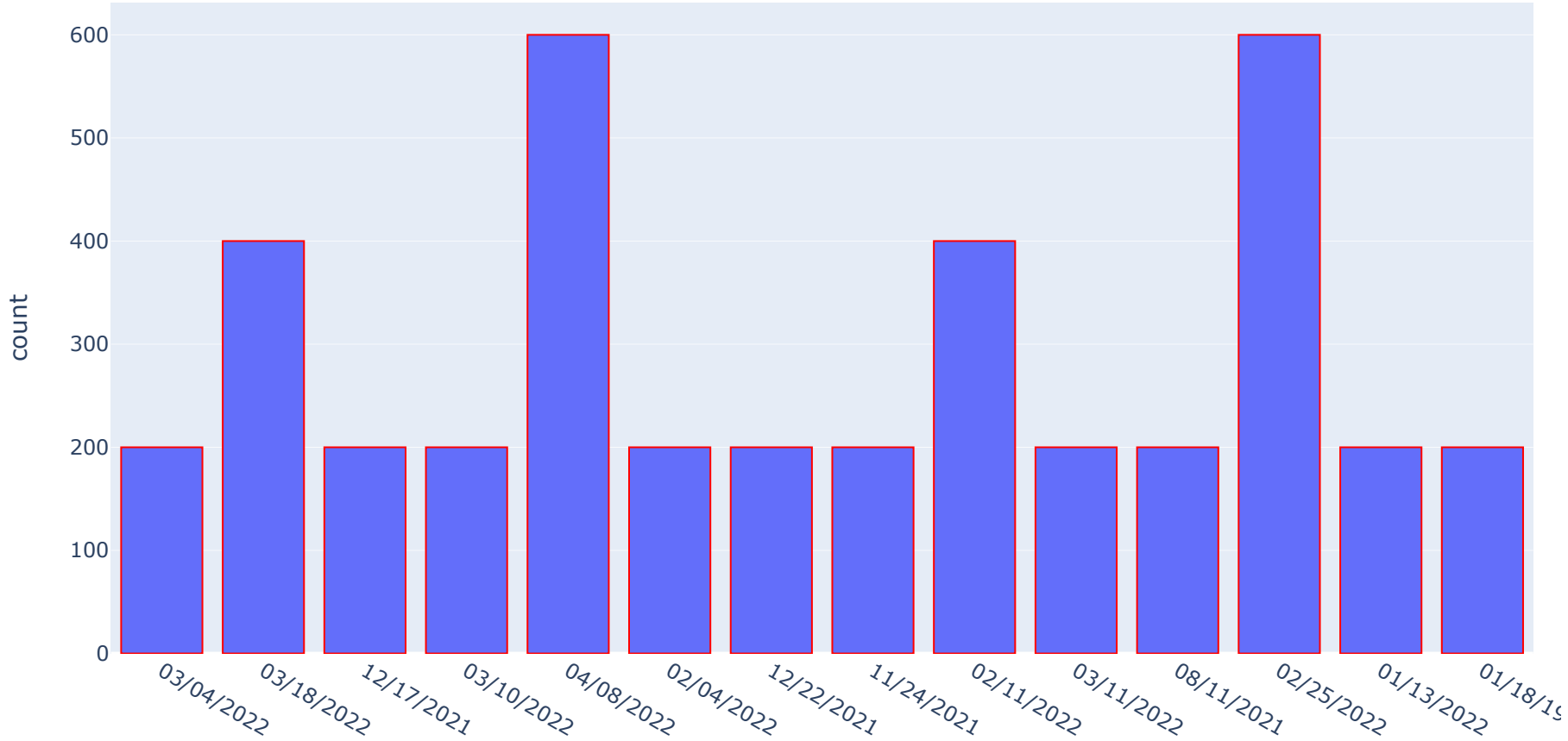
Out[214]: <AxesSubplot:xlabel='Rating', ylabel='Density'>



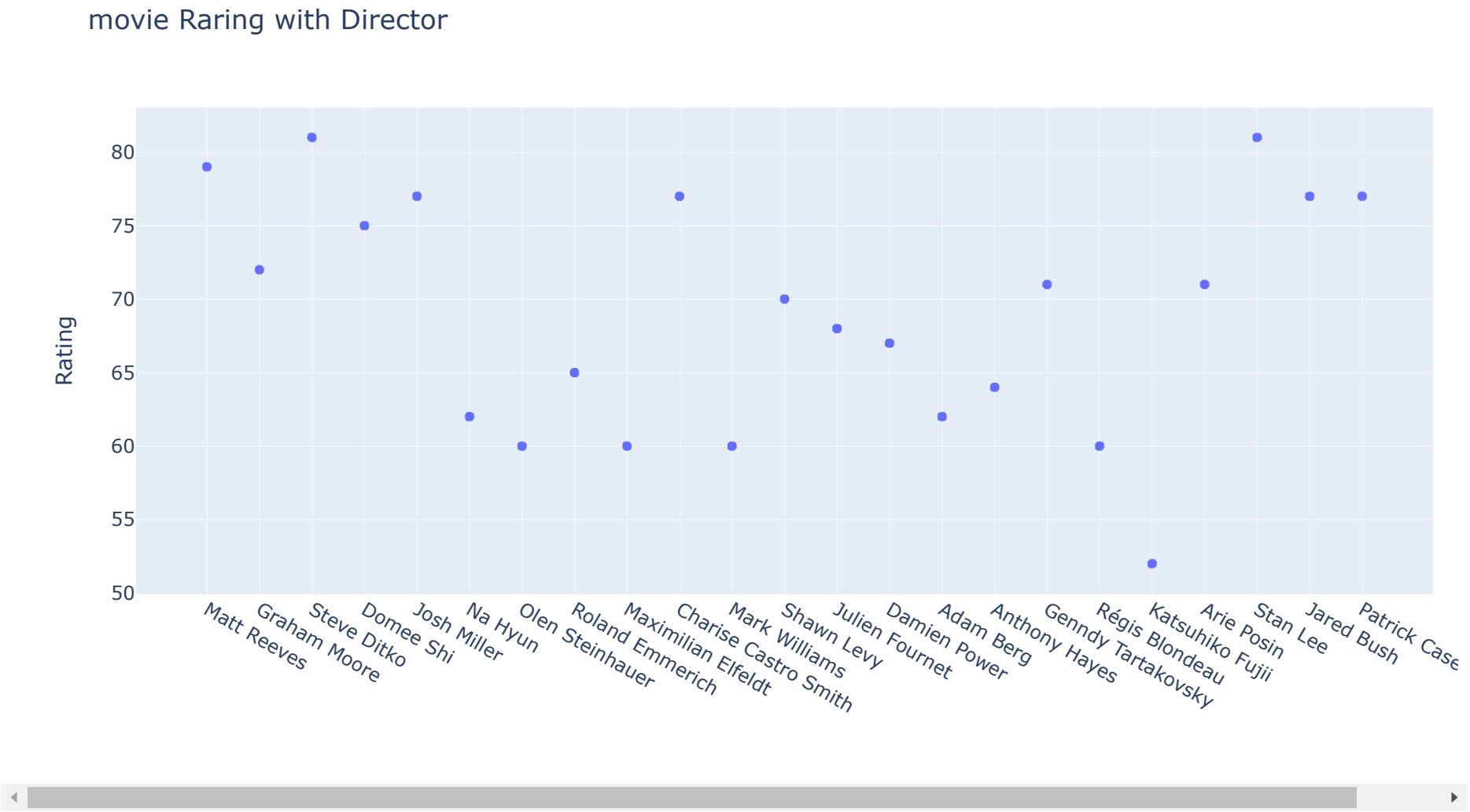
In [174]:
```
px.scatter(x='Rating',y='Name',data_frame=data,color_discrete_sequence=px.colors.qualitative.Plotly_r)
```



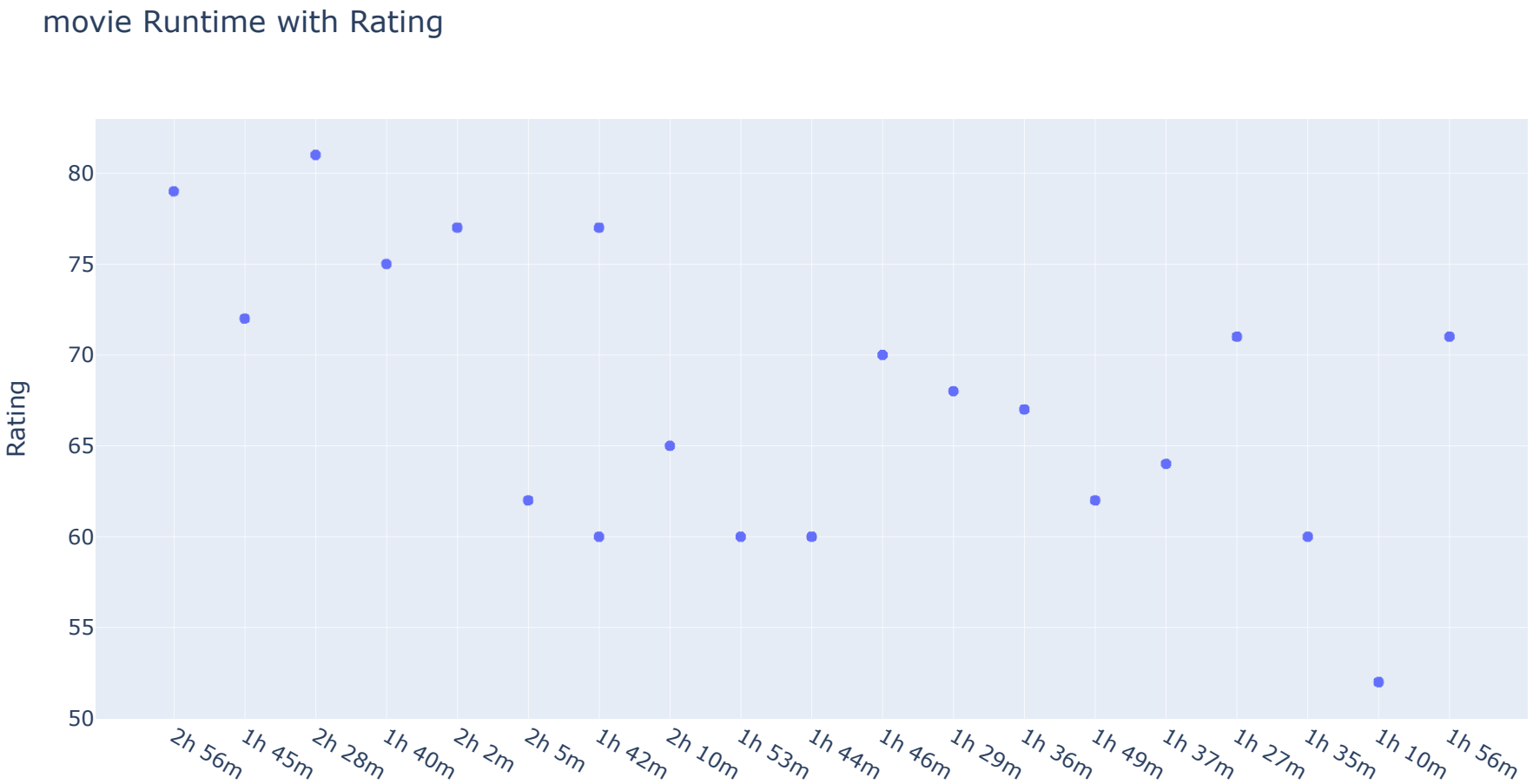
In [23]:
```
fig = px.histogram(x='Date',data_frame=data)
fig.update_traces(marker_line_width=1,marker_line_color="red")
```



In [24]:
```
px.scatter(x='Director',y='Rating',data_frame=data,title='movie Raring with Director')
```

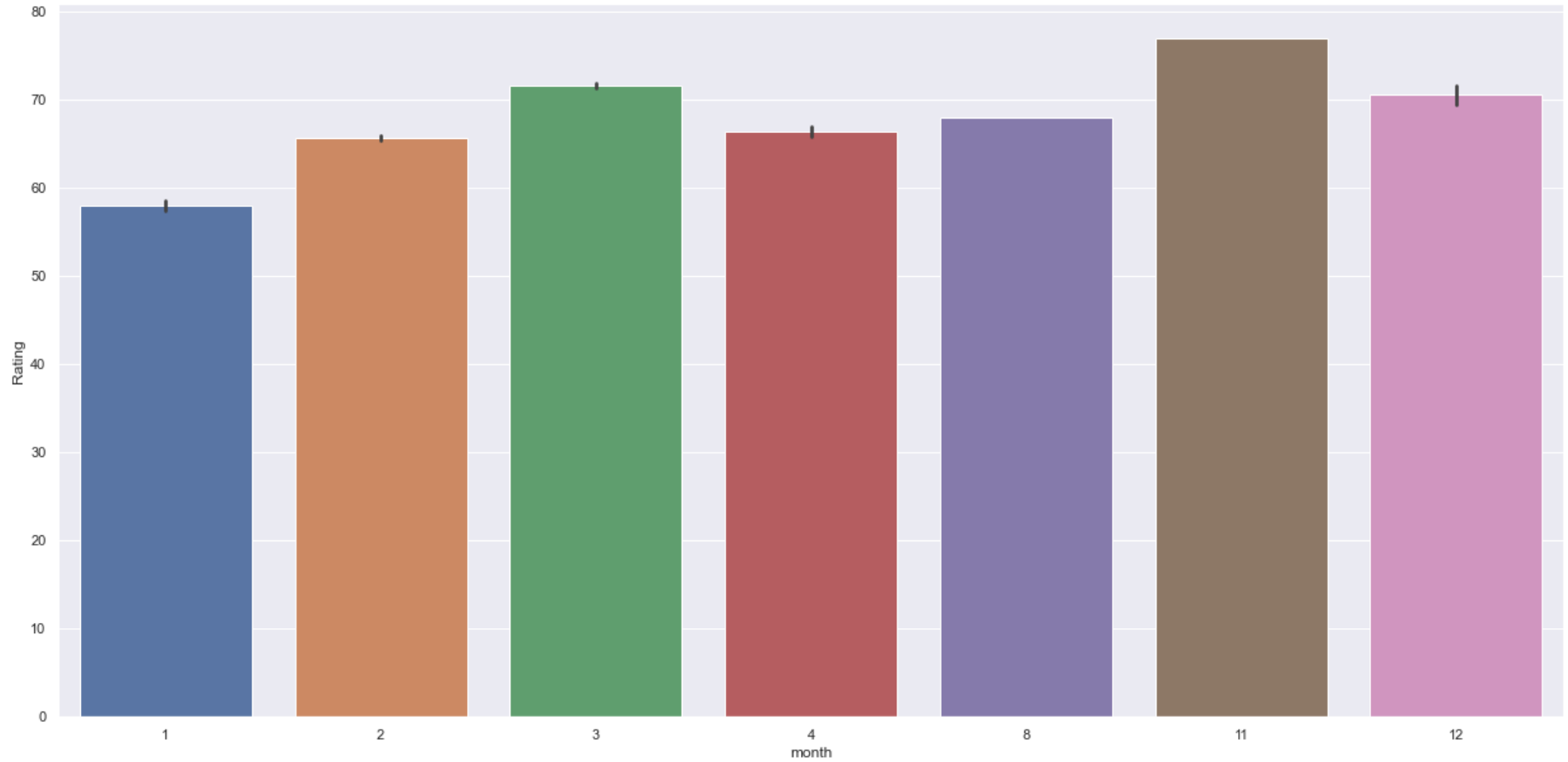


In [26]:
```
px.scatter(x='Runtime',y='Rating',data_frame=data,title='movie Runtime with Rating')
```



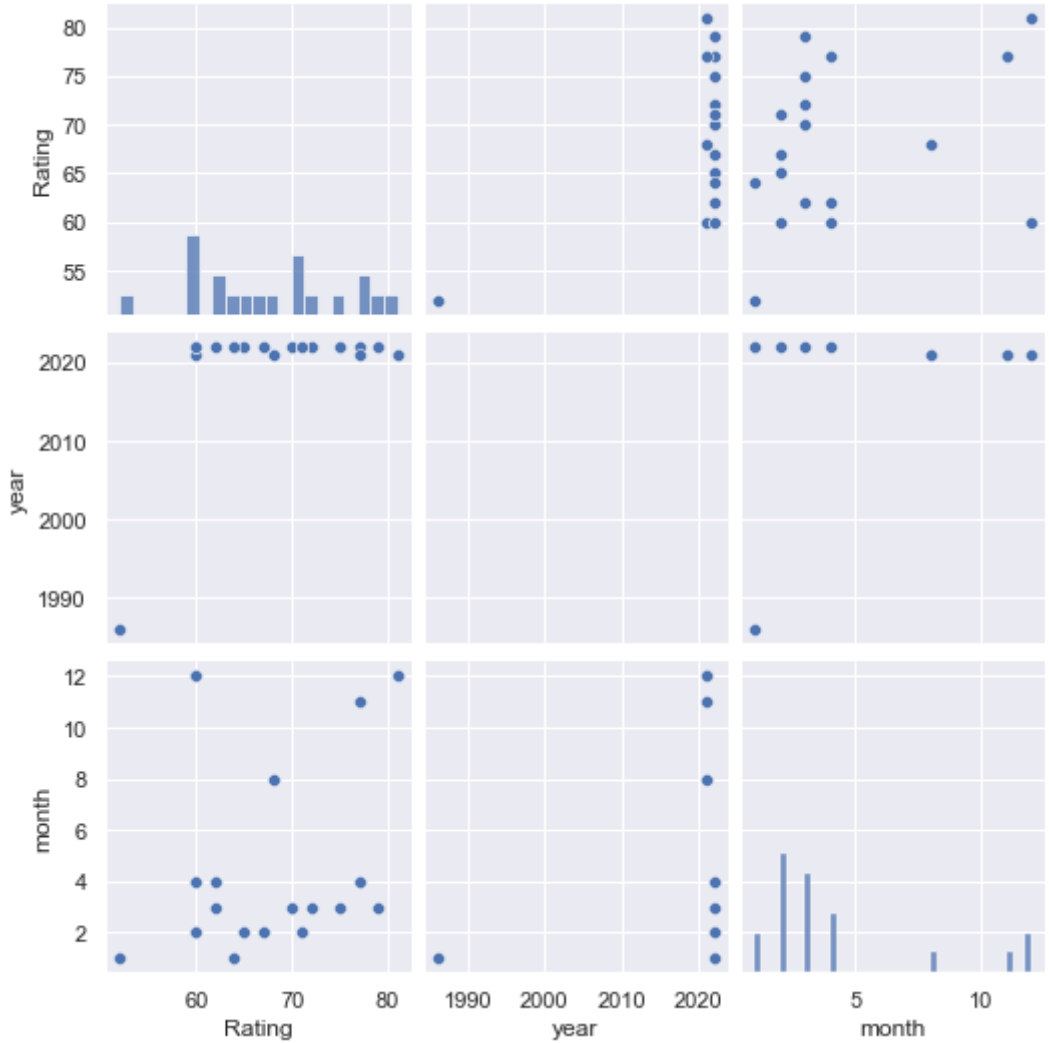
In [177]:
```
sns.barplot(x='month',y='Rating',data=data)
```

Out[177]: <AxesSubplot:xlabel='month', ylabel='Rating'>



In [171]:
```
sns.pairplot(data)
```

Out[171]:
```
<seaborn.axisgrid.PairGrid at 0x241463aceb0>
```

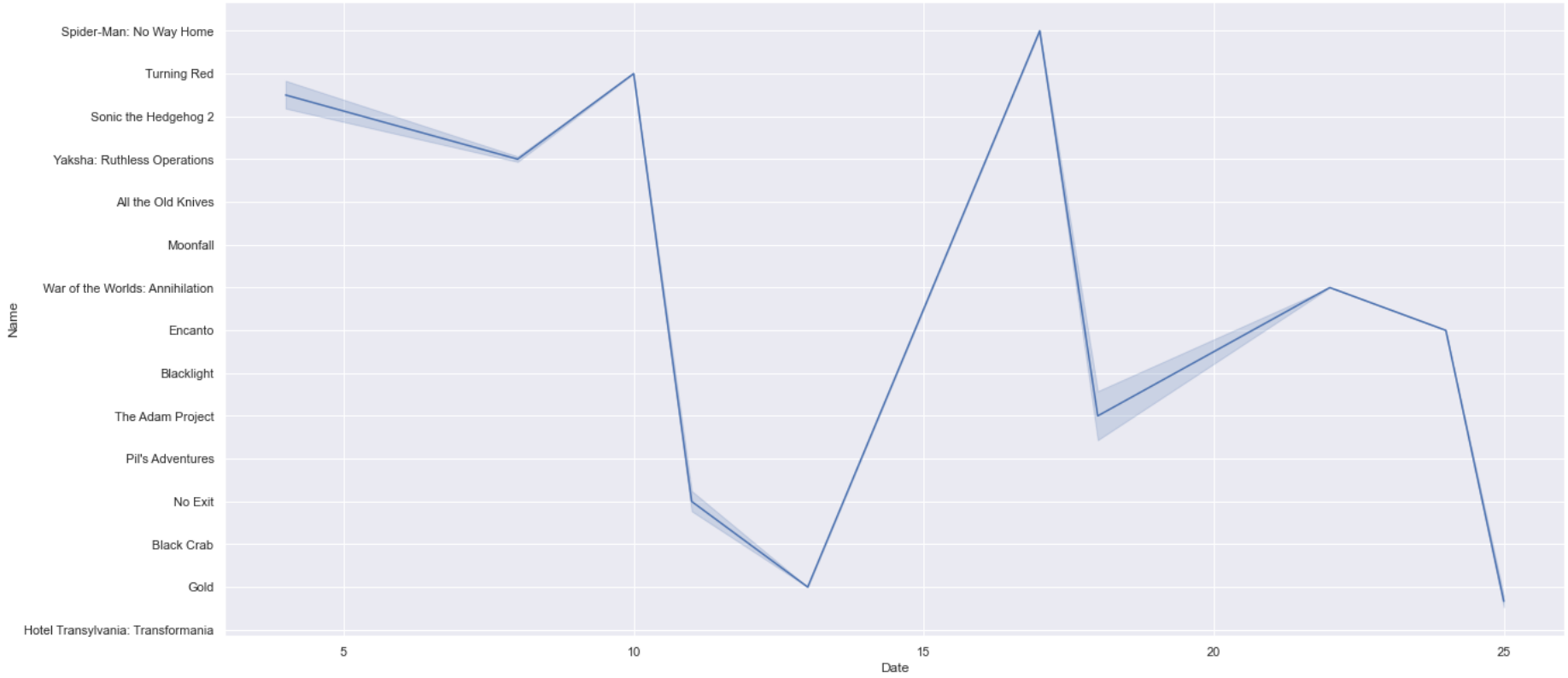


In [120]:
```
#data['Date']=pd.to_datatime(data['Date'],format="%m/%d/%y").Date
#date['Year']=pd.to_DatetimeIndex(data['Date']).Year
#date['Month']=pd.to_DatetimeIndex(data['Month']).Month
#date['Day']=pd.to_DatefrmaeIndex(data['Day']).Day
```

Year

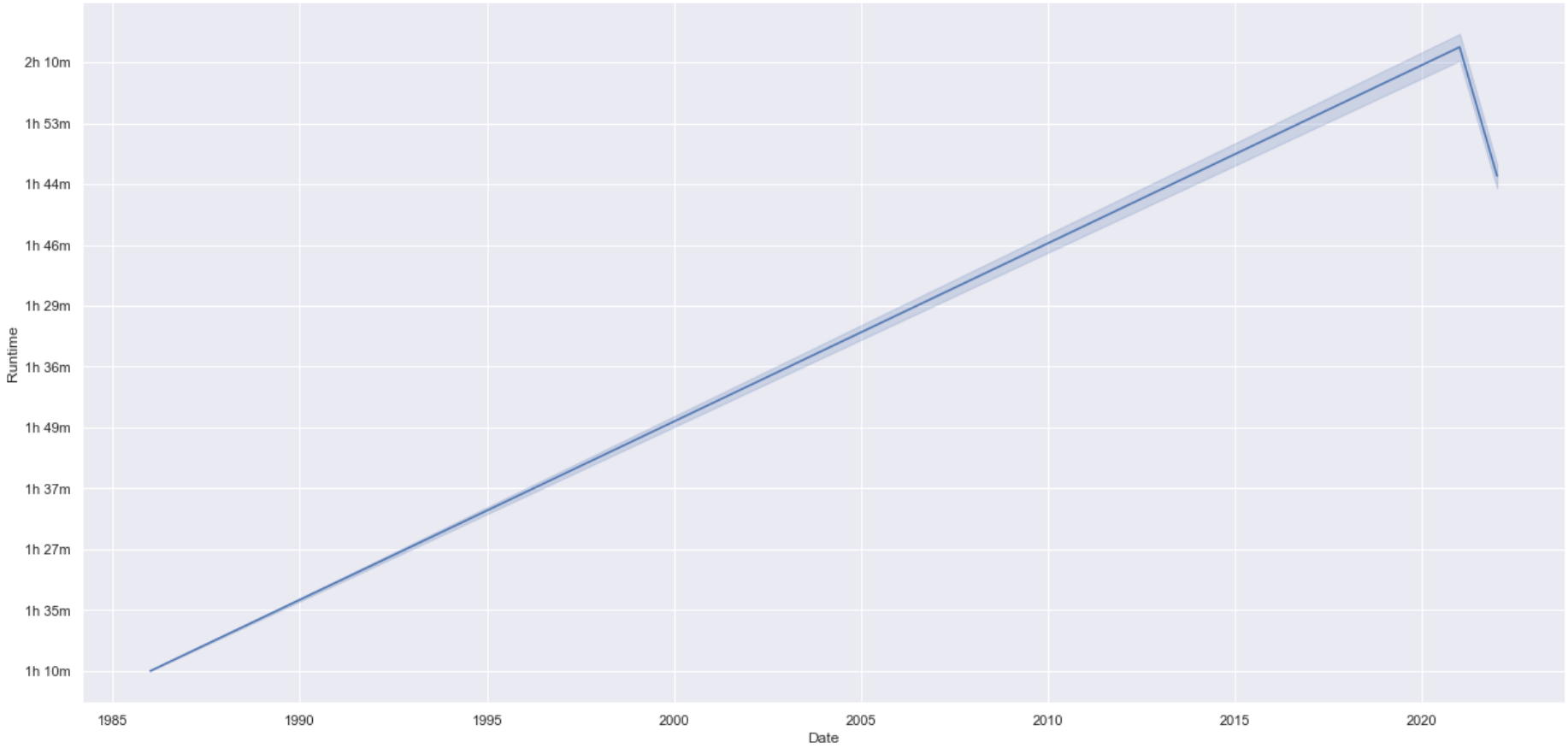
In [183]:
```
sns.lineplot(data=data, x=pd.DatetimeIndex(data['Date']).year, y="Name")
```

Out[183]:
```
<AxesSubplot:xlabel='Date', ylabel='Name'>
```



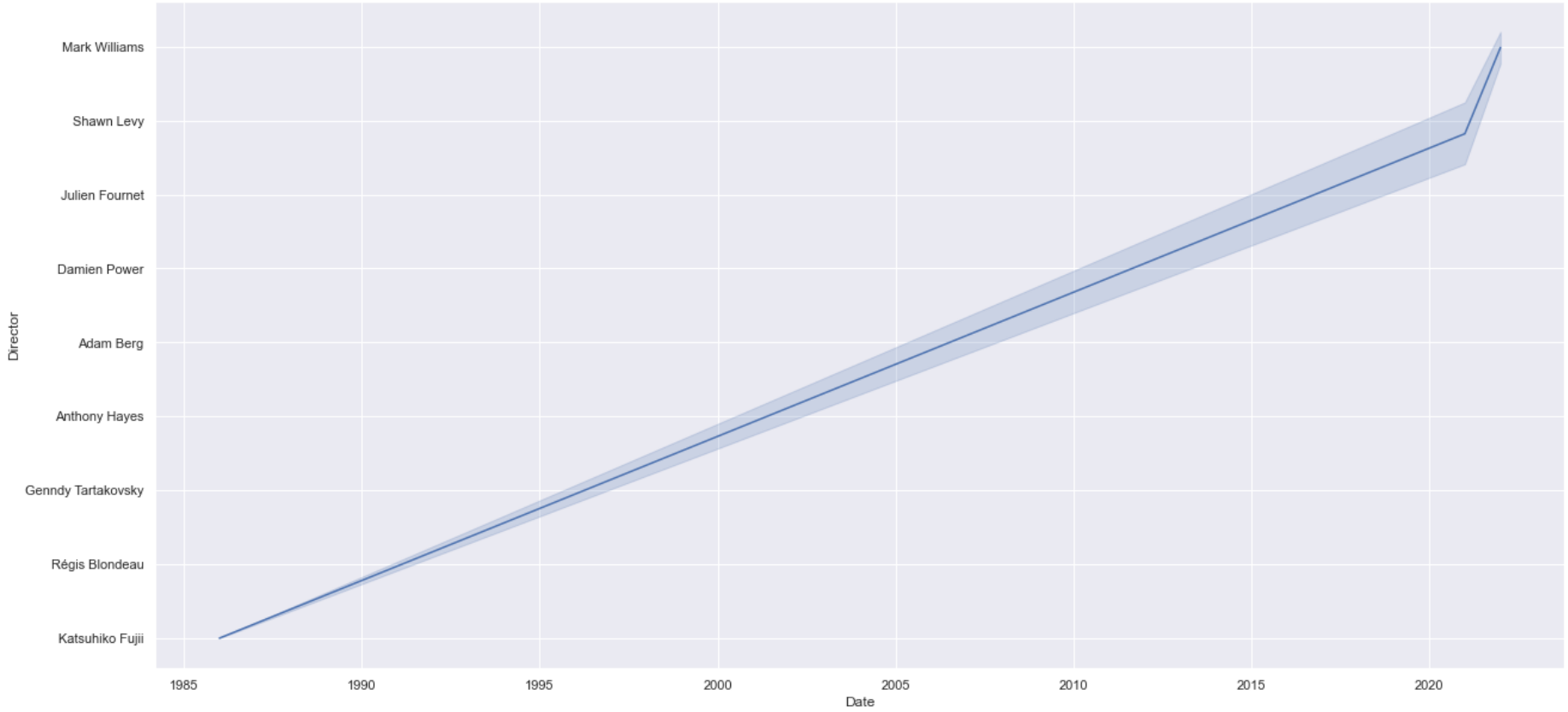
In [187]:
```
sns.lineplot(data=data, x=pd.DatetimeIndex(data['Date']).year, y="Runtime")
```

Out[187]:
```
<AxesSubplot:xlabel='Date', ylabel='Runtime'>
```



In [188]:
```
sns.lineplot(data=data, x=pd.DatetimeIndex(data['Date']).year, y="Director")
```

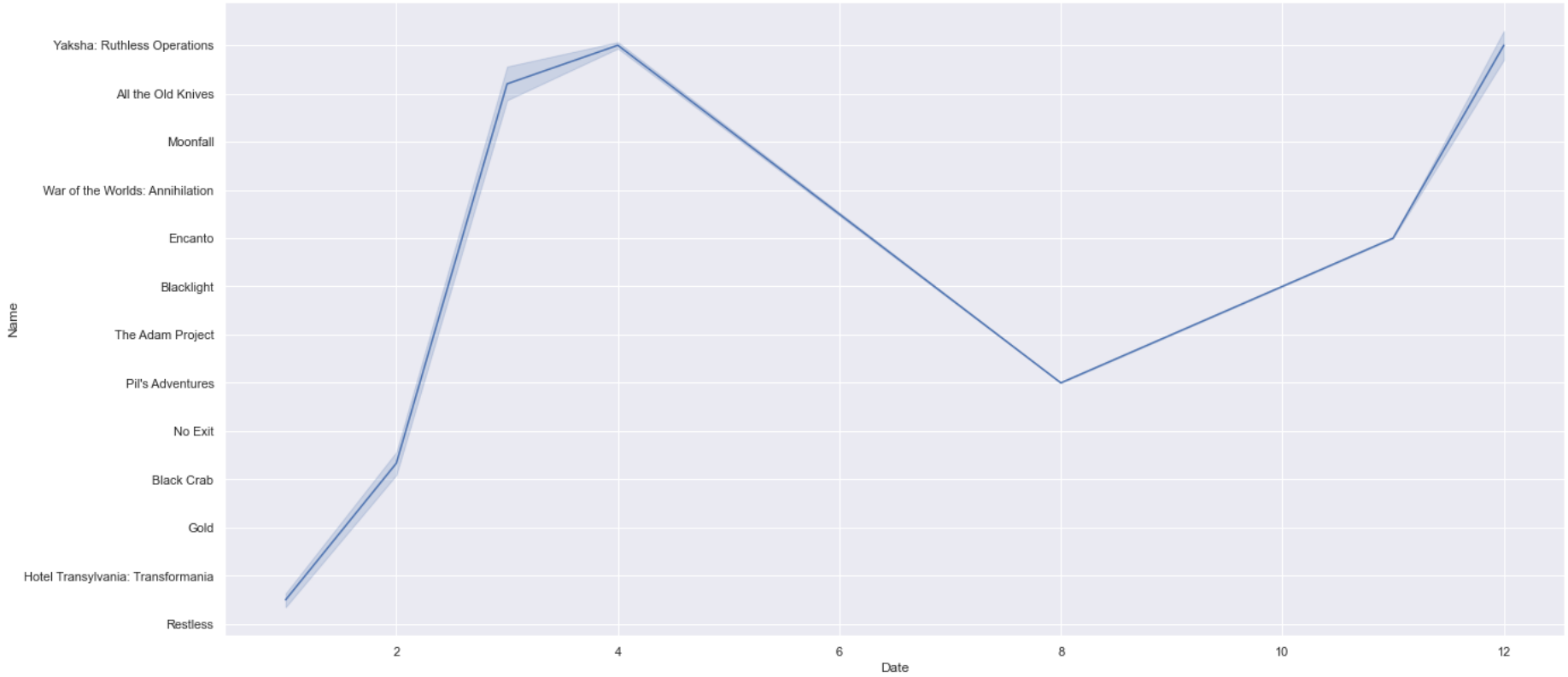
Out[188]:
```
<AxesSubplot:xlabel='Date', ylabel='Director'>
```



Month

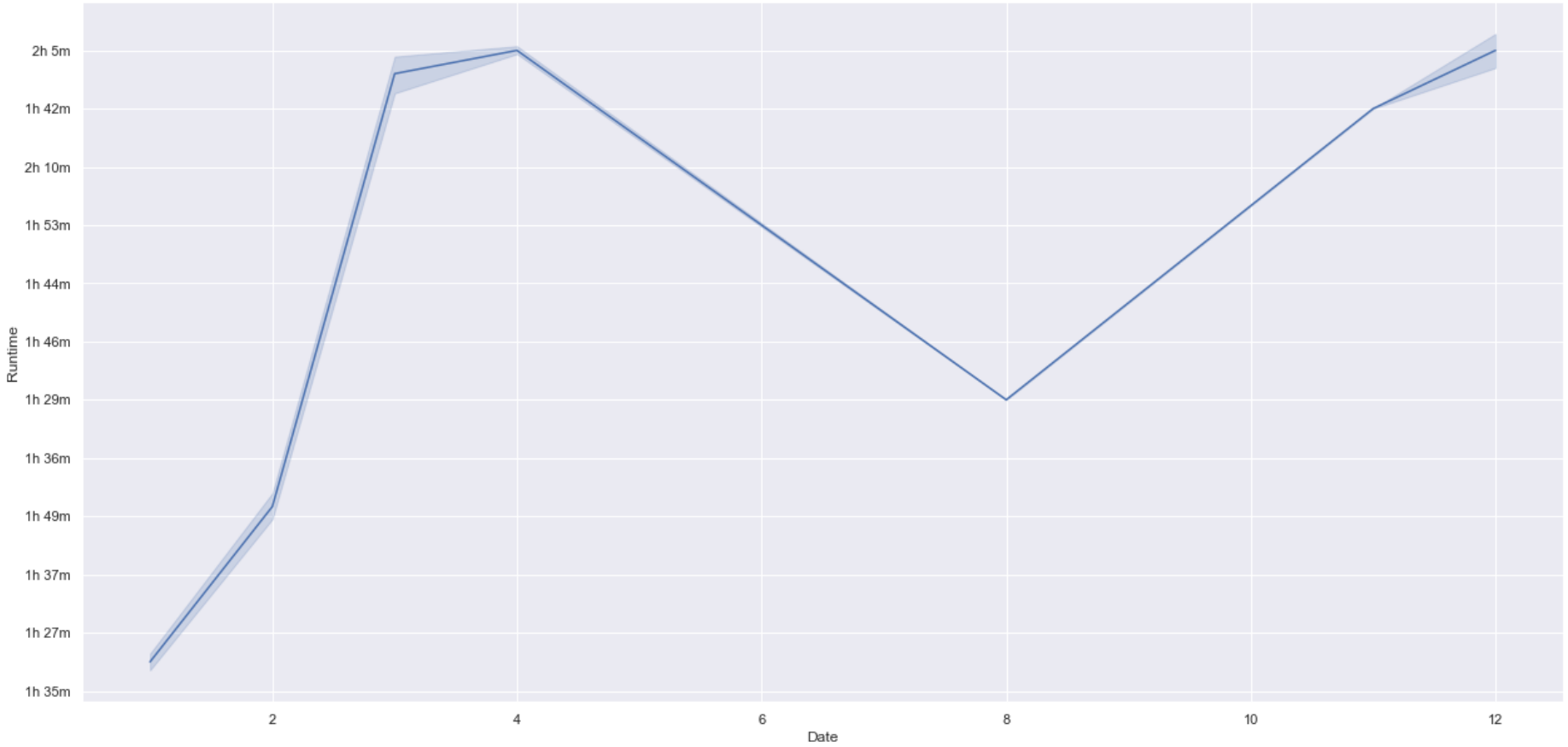
In [189]:
```
sns.lineplot(data=data, x=pd.DatetimeIndex(data['Date']).month, y="Name")
```

Out[189]:
```
<AxesSubplot:xlabel='Date', ylabel='Name'>
```



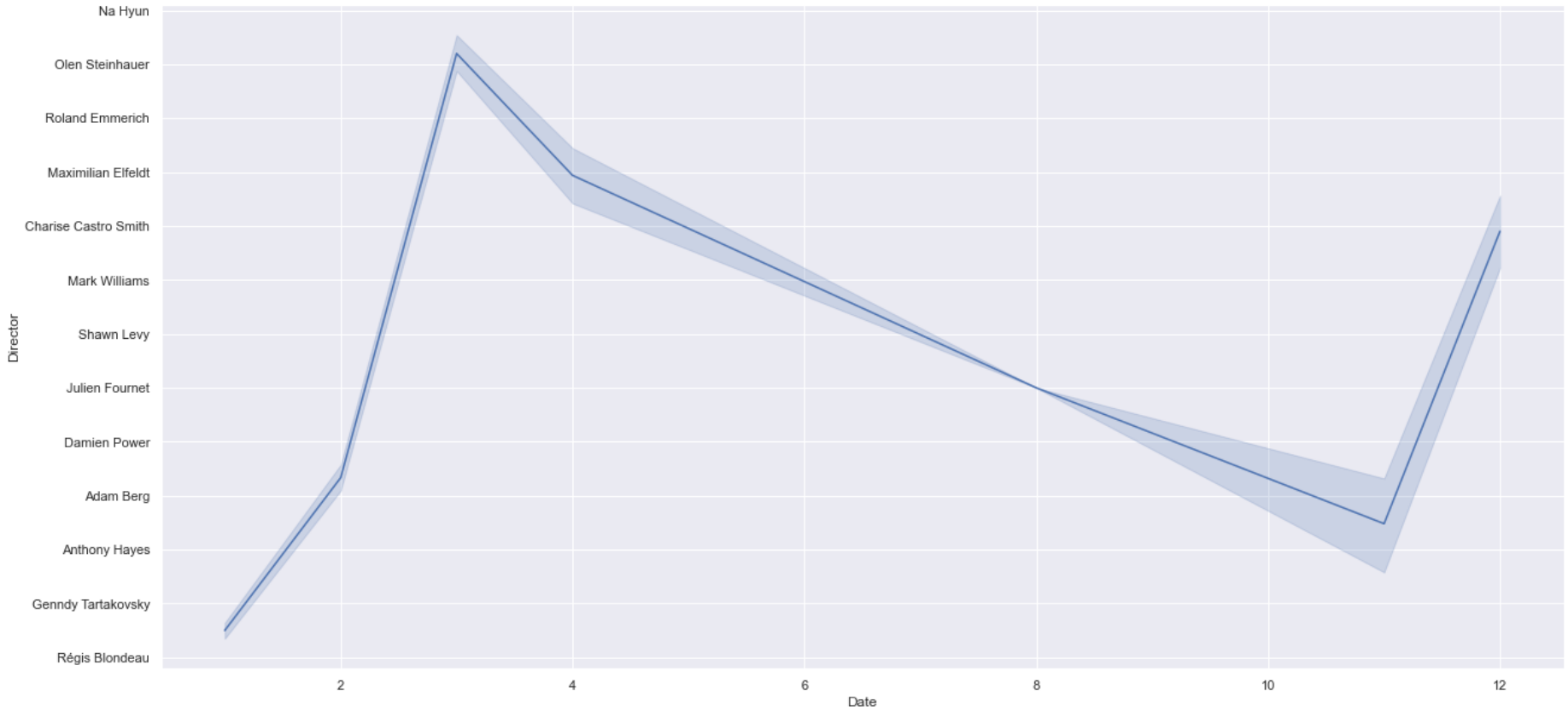
In [191]:
```
sns.lineplot(data=data, x=pd.DatetimeIndex(data['Date']).month, y="Runtime")
```

Out[191]:
```
<AxesSubplot:xlabel='Date', ylabel='Runtime'>
```



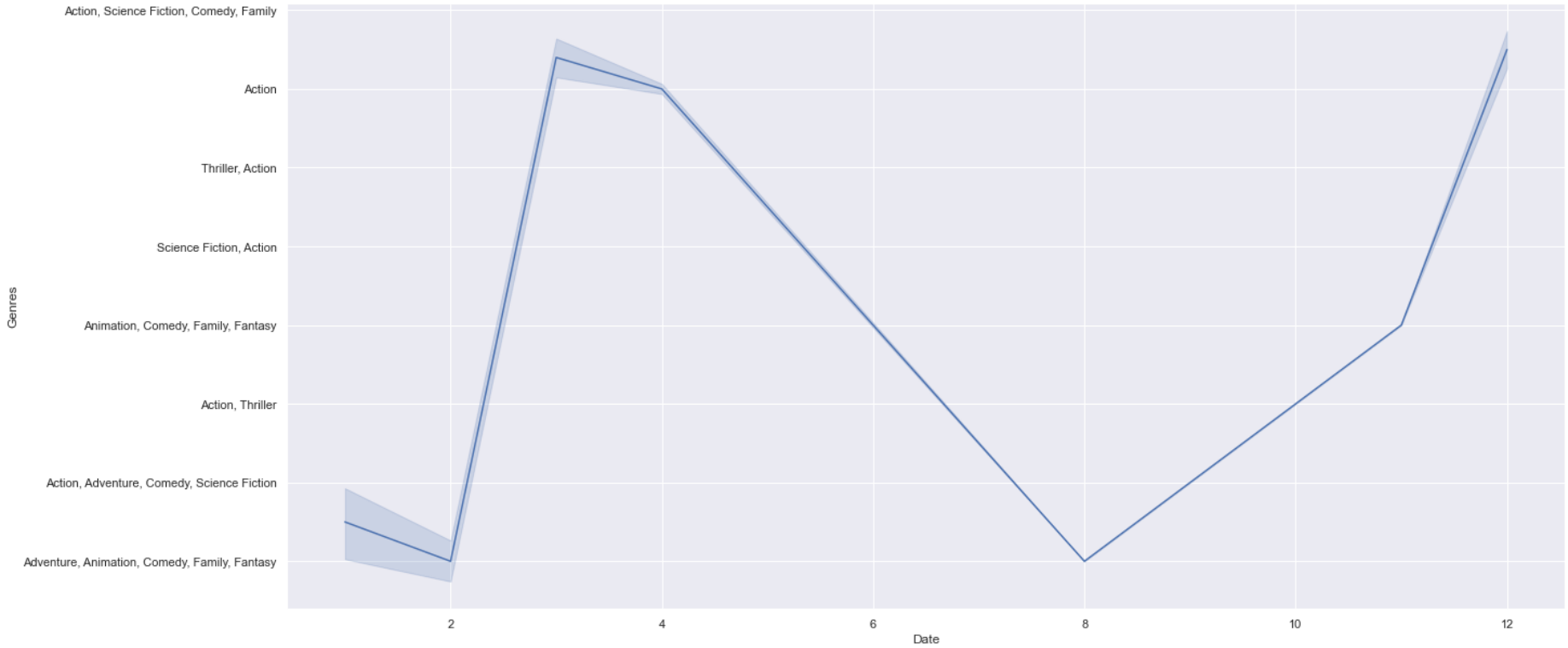
In [192]:
```
sns.lineplot(data=data, x=pd.DatetimeIndex(data['Date']).month, y="Director")
```

Out[192]: <AxesSubplot:xlabel='Date', ylabel='Director'>



In [193]:
```
sns.lineplot(data=data, x=pd.DatetimeIndex(data['Date']).month, y="Genres")
```

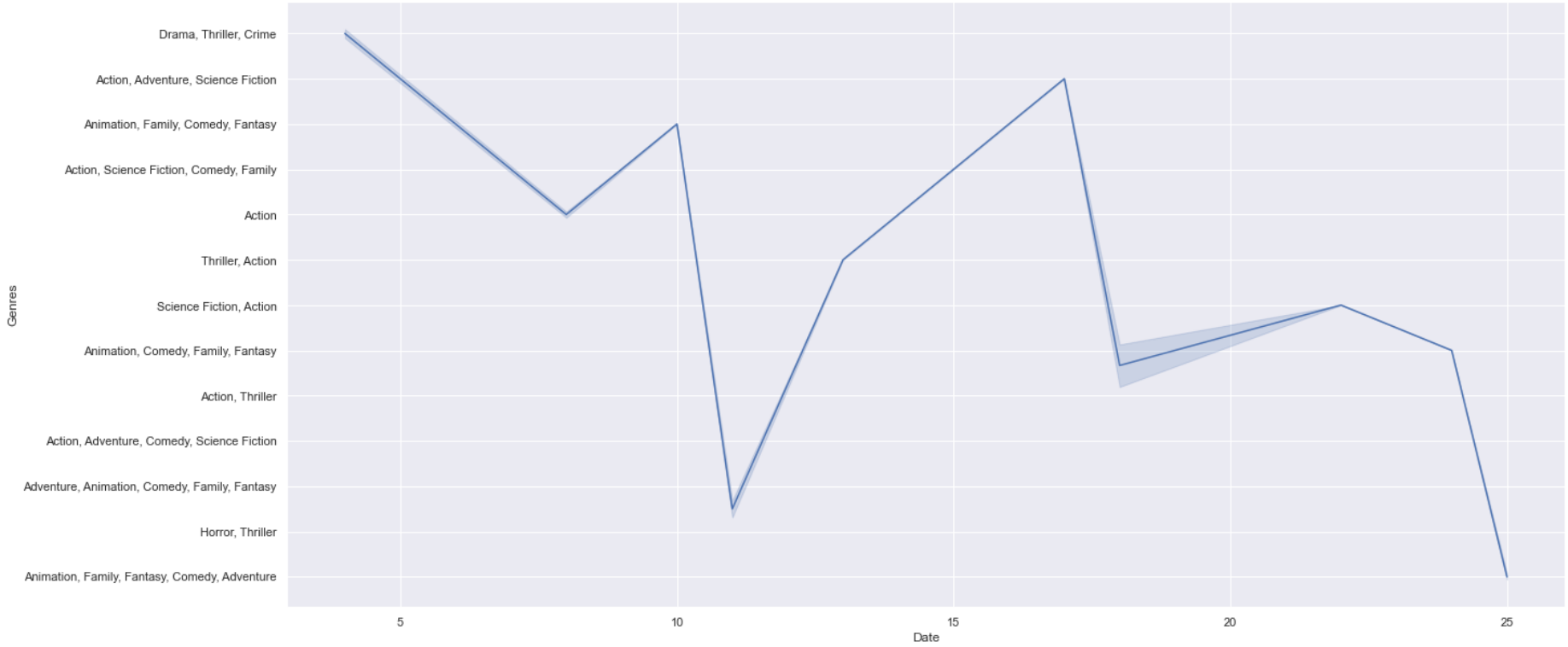
Out[193]:
```
<AxesSubplot:xlabel='Date', ylabel='Genres'>
```



Day

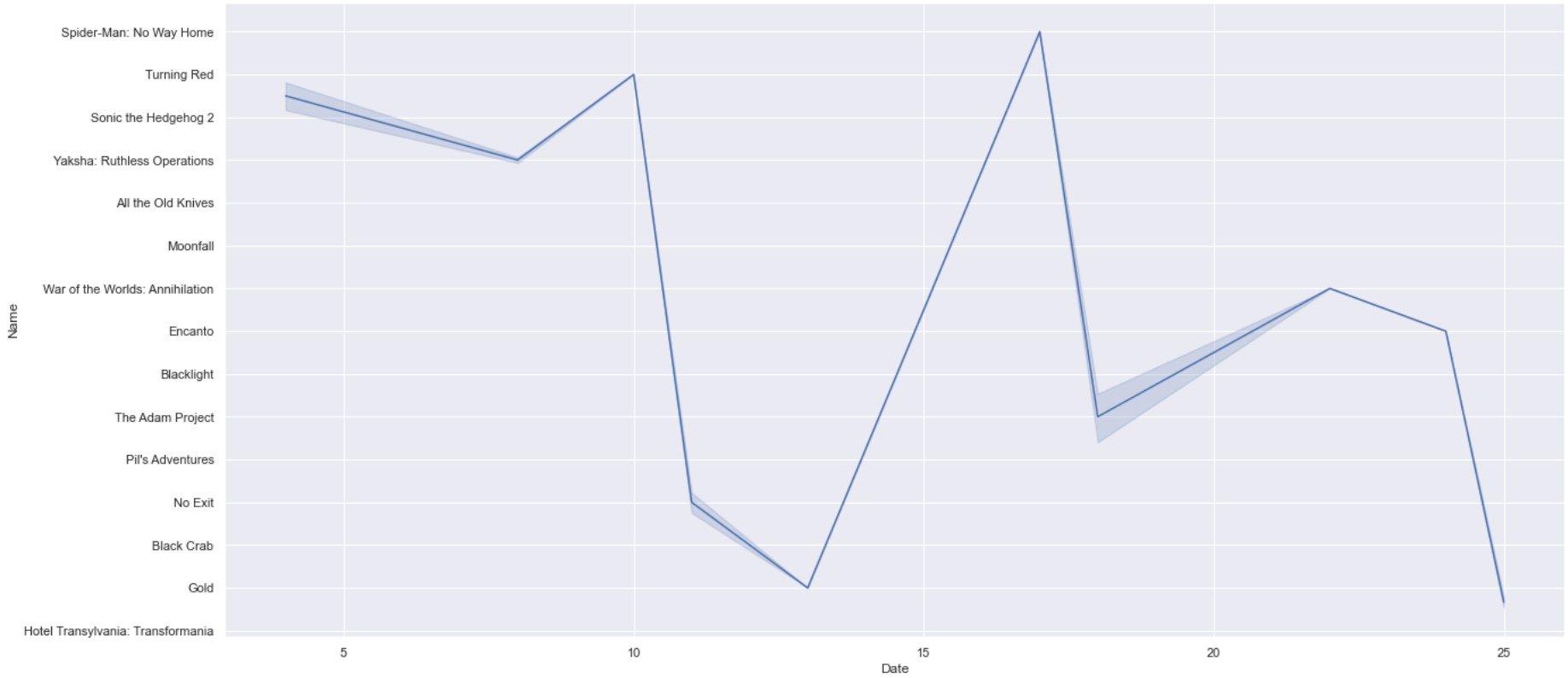
In [205]:
```
sns.lineplot(data=data, x=pd.DatetimeIndex(data['Date']).day, y="Genres")
```

Out[205]: <AxesSubplot:xlabel='Date', ylabel='Genres'>



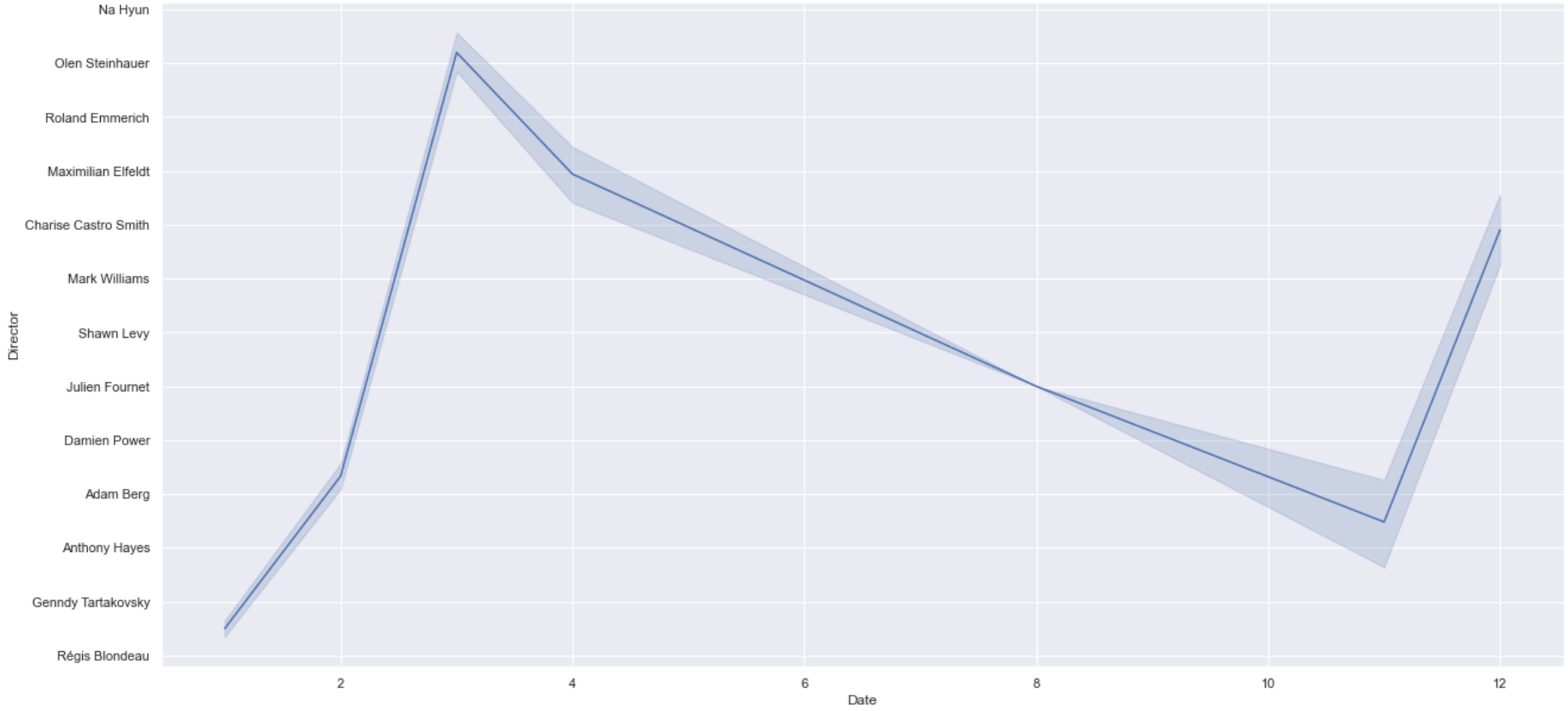
In [206]:
```
sns.lineplot(data=data, x=pd.DatetimeIndex(data['Date']).day, y="Name")
```

Out[206]:
```
<AxesSubplot:xlabel='Date', ylabel='Name'>
```



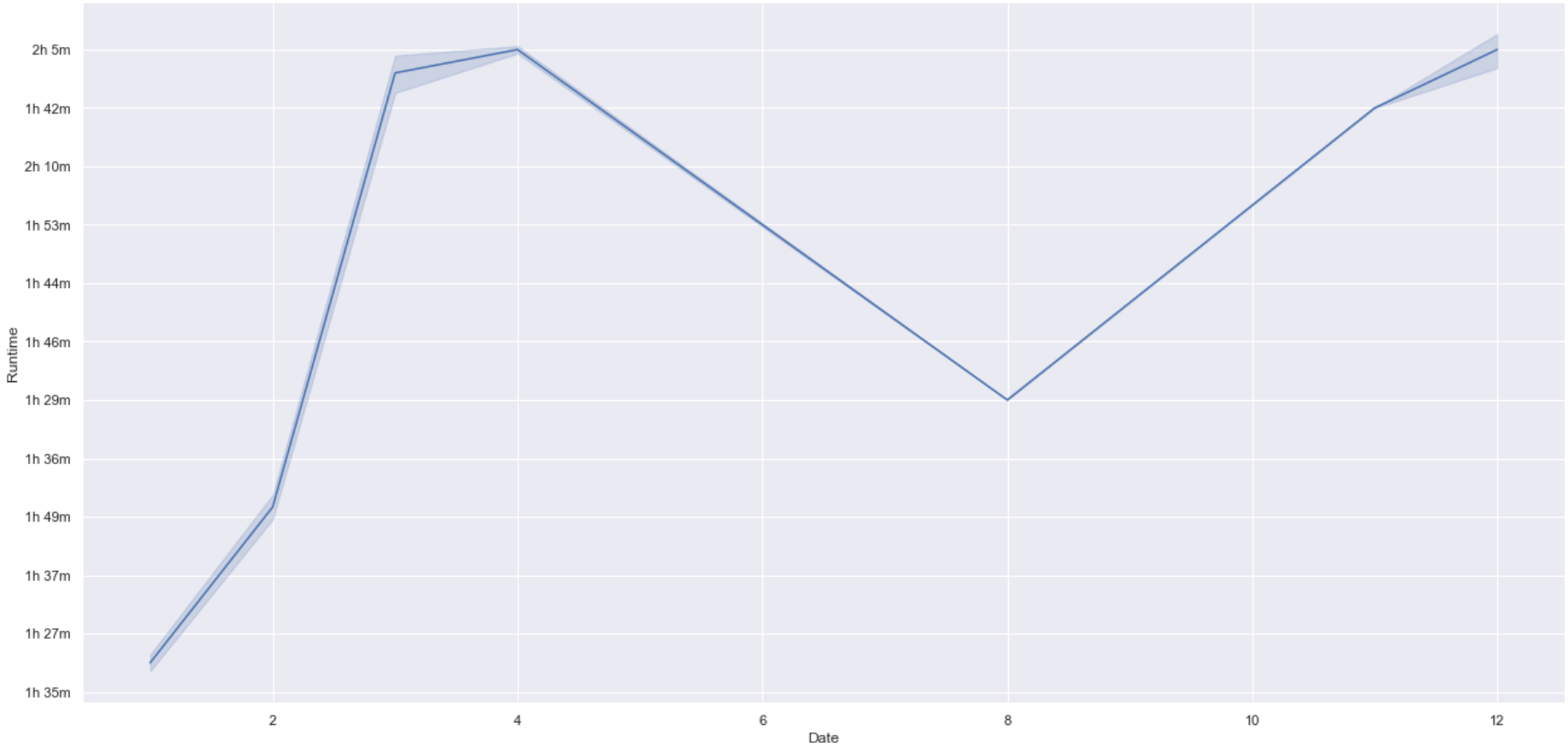
In [207]:
```
sns.lineplot(data=data, x=pd.DatetimeIndex(data['Date']).month, y="Director")
```

Out[207]: <AxesSubplot:xlabel='Date', ylabel='Director'>



In [208]:
```
sns.lineplot(data=data, x=pd.DatetimeIndex(data['Date']).month, y="Runtime")
```

Out[208]:
```
<AxesSubplot:xlabel='Date', ylabel='Runtime'>
```



In [ ]:

In [201]:
```
df1=data.groupby(['Name'])[['Rating']].max()
df1
```

Out[201]:

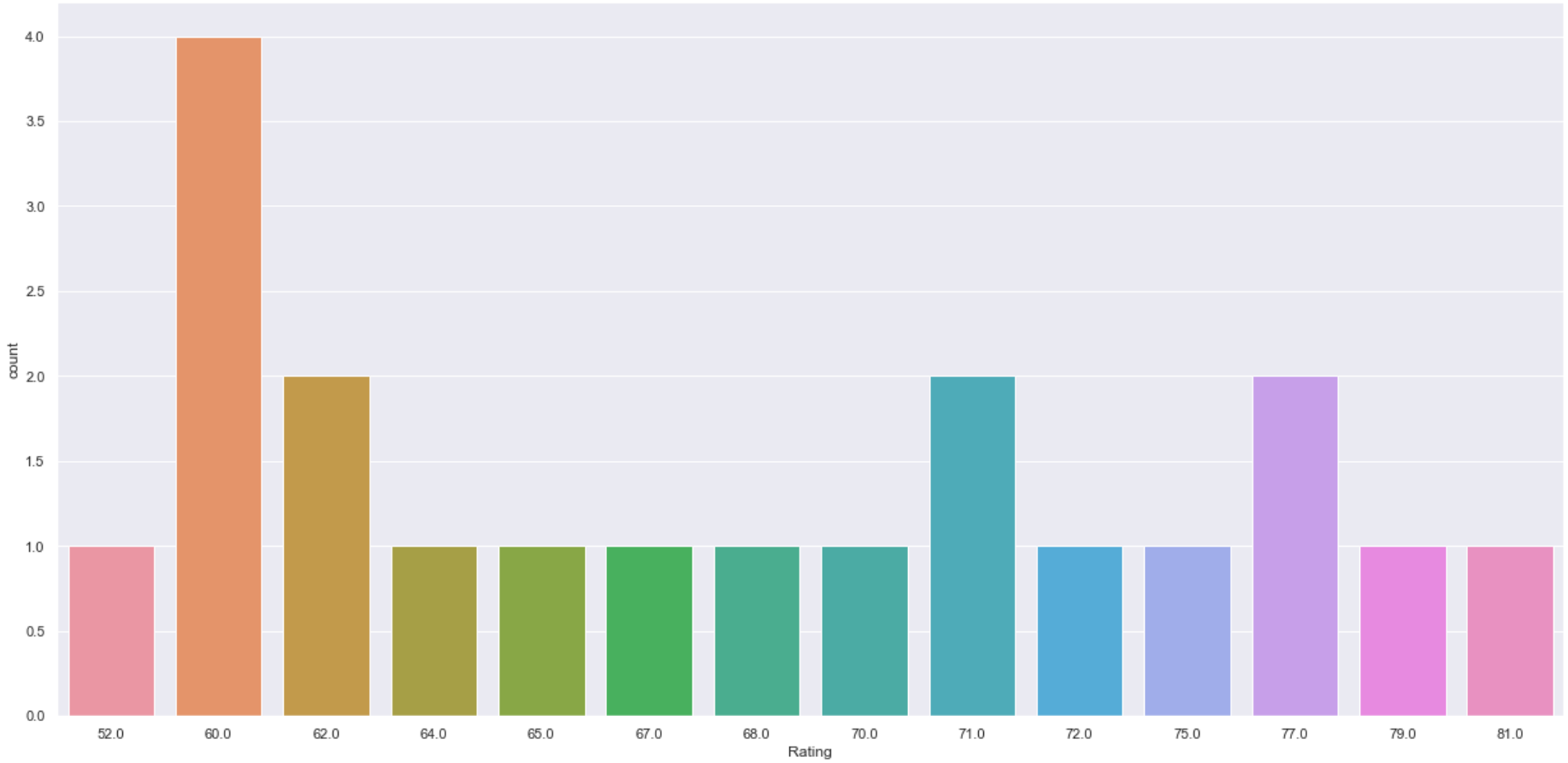
Name	Rating
All the Old Knives	60.0
Beautiful Sisters: Flesh Slave	52.0
Black Crab	62.0
Blacklight	60.0
Encanto	77.0
Gold	64.0
Hotel Transylvania: Transformania	71.0
Moonfall	65.0
No Exit	67.0
Pil's Adventures	68.0
Restless	60.0
Sonic the Hedgehog 2	77.0
Spider-Man: No Way Home	81.0
The Adam Project	70.0
The Batman	79.0
The In Between	71.0
The Outfit	72.0
Turning Red	75.0
War of the Worlds: Annihilation	60.0
Yaksha: Ruthless Operations	62.0

In [203]:
```
sns.countplot(df1['Rating'])
```

```
C:\Users\Admin\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning:

Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`,
and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

Out[203]:
```
<AxesSubplot:xlabel='Rating', ylabel='count'>
```



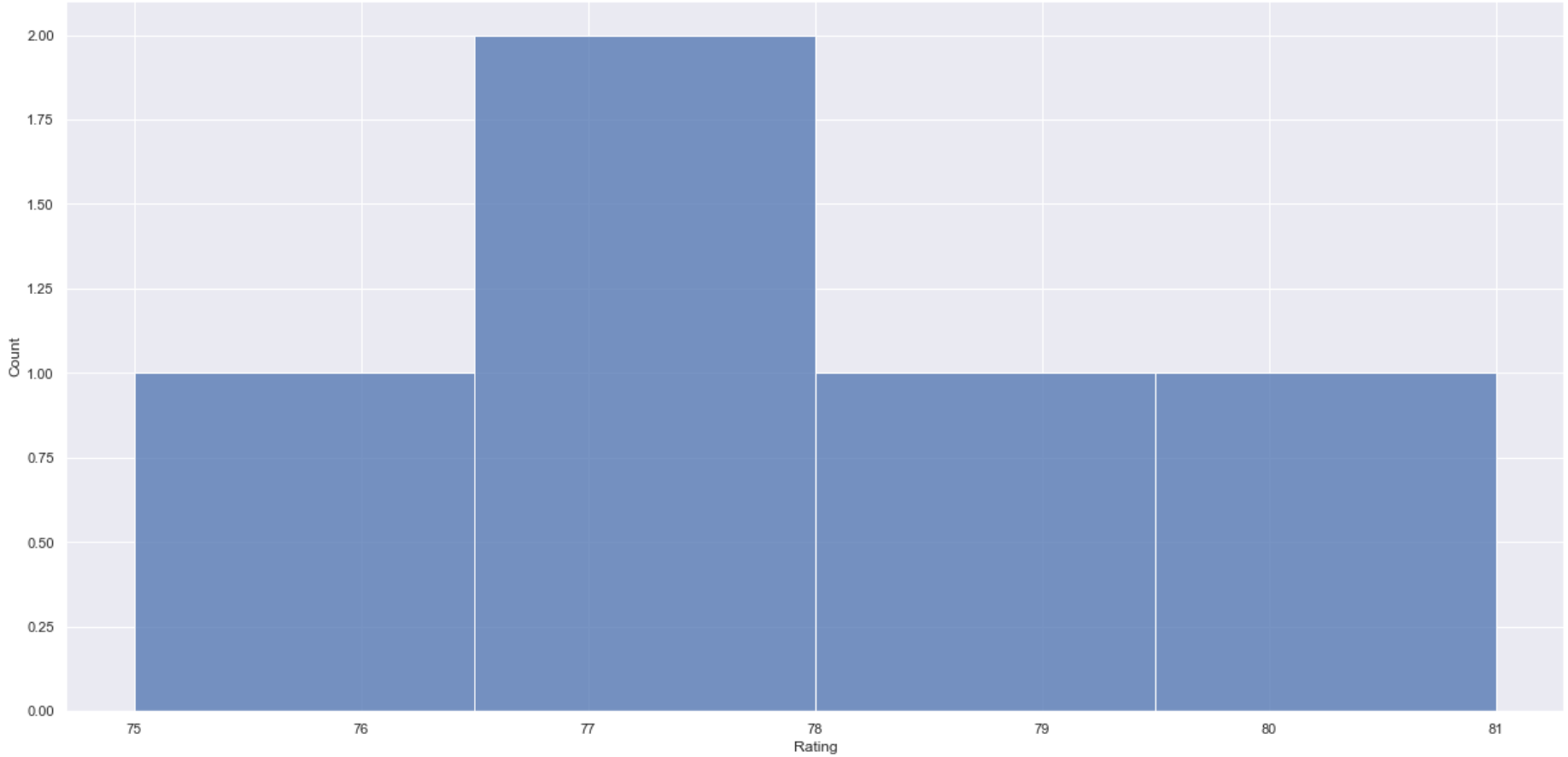
In [202]:
```
s1 = df1.sort_values(by='Rating',ascending=False).head()
s1
```

Out[202]:

Name	Rating
Spider-Man: No Way Home	81.0
The Batman	79.0
Encanto	77.0
Sonic the Hedgehog 2	77.0
Turning Red	75.0

In [204]:
```
sns.histplot(s1['Rating'])
```

Out[204]:
```
<AxesSubplot:xlabel='Rating', ylabel='Count'>
```



In [195]:
```
df2=data.groupby(['Director'])[['Runtime']].max()
df2
```

Out[195]:

Director	Runtime
Adam Berg	1h 49m
Anthony Hayes	1h 37m
Arie Posin	1h 56m
Charise Castro Smith	1h 42m
Damien Power	1h 36m
Domee Shi	1h 40m
Genndy Tartakovsky	1h 27m
Graham Moore	1h 45m
Jared Bush	1h 42m
Josh Miller	2h 2m
Julien Fournet	1h 29m
Katsuhiko Fujii	1h 10m
Mark Williams	1h 44m
Matt Reeves	2h 56m
Maximilian Elfeldt	1h 53m
Na Hyun	2h 5m
Olen Steinhauer	1h 42m
Patrick Casey	2h 2m
Roland Emmerich	2h 10m
Régis Blondeau	1h 35m
Shawn Levy	1h 46m
Stan Lee	2h 28m

Director	Runtime
Steve Ditko	2h 28m

In [196]:
```
s2=df2.sort_values(by='Runtime',ascending=False).head()
s2
```

Out[196]:

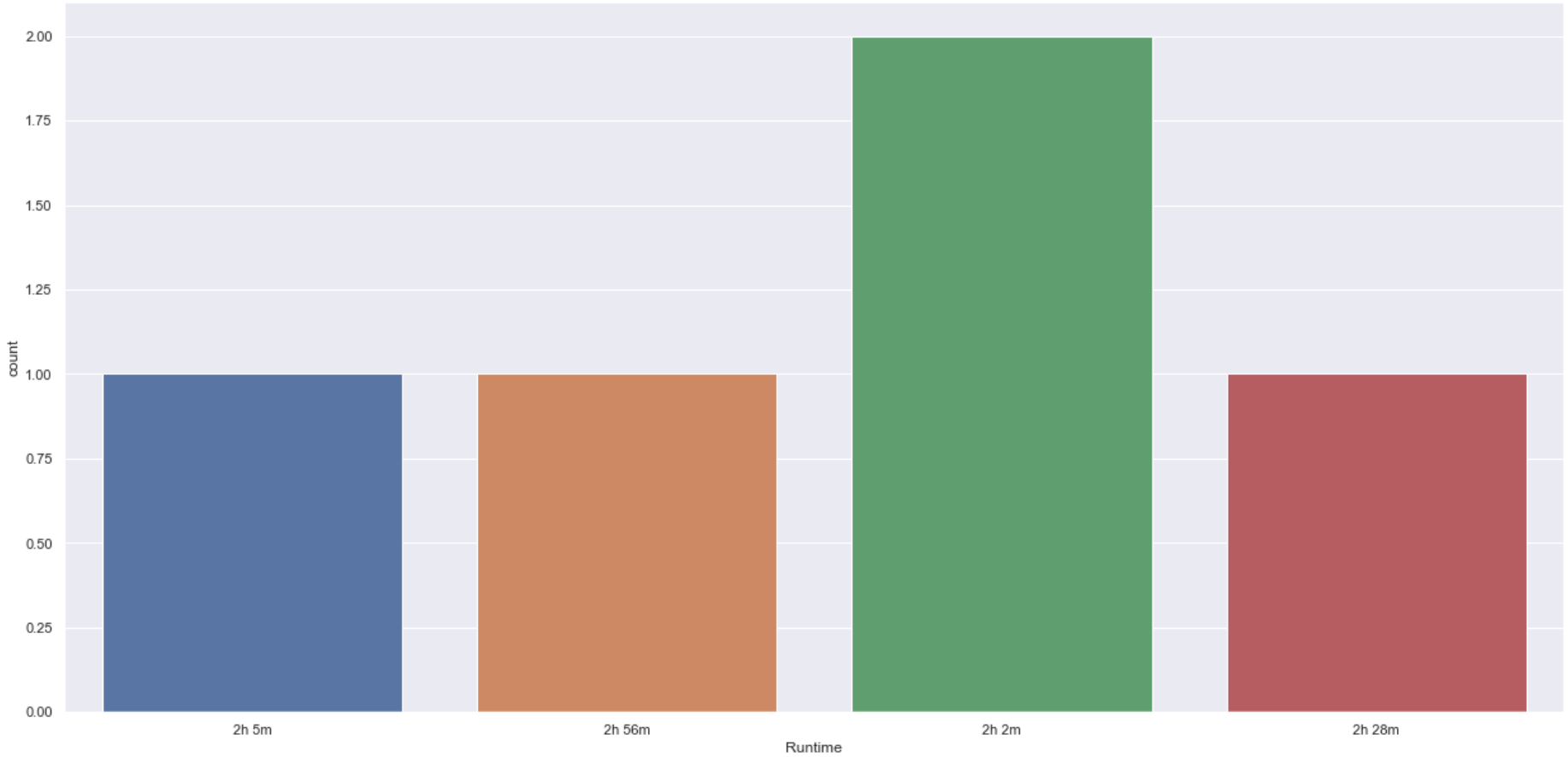
Director	Runtime
Na Hyun	2h 5m
Matt Reeves	2h 56m
Josh Miller	2h 2m
Patrick Casey	2h 2m
Steve Ditko	2h 28m

In [197]:
```
sns.countplot(s2['Runtime'])
```

```
C:\Users\Admin\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning:

Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`,
and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

Out[197]:
```
<AxesSubplot:xlabel='Runtime', ylabel='count'>
```



In [42]:
```
df3=data.groupby(['Runtime'])[['Rating']].max()
df3
```

Out[42]:

Runtime	Rating
1h 10m	52.0
1h 27m	71.0
1h 29m	68.0
1h 35m	60.0
1h 36m	67.0
1h 37m	64.0
1h 40m	75.0
1h 42m	77.0
1h 44m	60.0
1h 45m	72.0
1h 46m	70.0
1h 49m	62.0
1h 53m	60.0
1h 56m	71.0
2h 10m	65.0
2h 28m	81.0
2h 2m	77.0
2h 56m	79.0
2h 5m	62.0

In [43]:
```
s3=df3.sort_values(by='Rating',ascending=False).head()
s3
```

Out[43]:

Runtime	Rating
2h 28m	81.0
2h 56m	79.0
1h 42m	77.0
2h 2m	77.0
1h 40m	75.0

In [ ]:

In [44]:
```
df4=data.groupby(['Name'])[['Runtime']].max()
df4
```

Out[44]:

Name	Runtime
All the Old Knives	1h 42m
Beautiful Sisters: Flesh Slave	1h 10m
Black Crab	1h 49m
Blacklight	1h 44m
Encanto	1h 42m
Gold	1h 37m
Hotel Transylvania: Transformania	1h 27m
Moonfall	2h 10m
No Exit	1h 36m
Pil's Adventures	1h 29m
Restless	1h 35m
Sonic the Hedgehog 2	2h 2m
Spider-Man: No Way Home	2h 28m
The Adam Project	1h 46m
The Batman	2h 56m
The In Between	1h 56m
The Outfit	1h 45m
Turning Red	1h 40m
War of the Worlds: Annihilation	1h 53m
Yaksha: Ruthless Operations	2h 5m

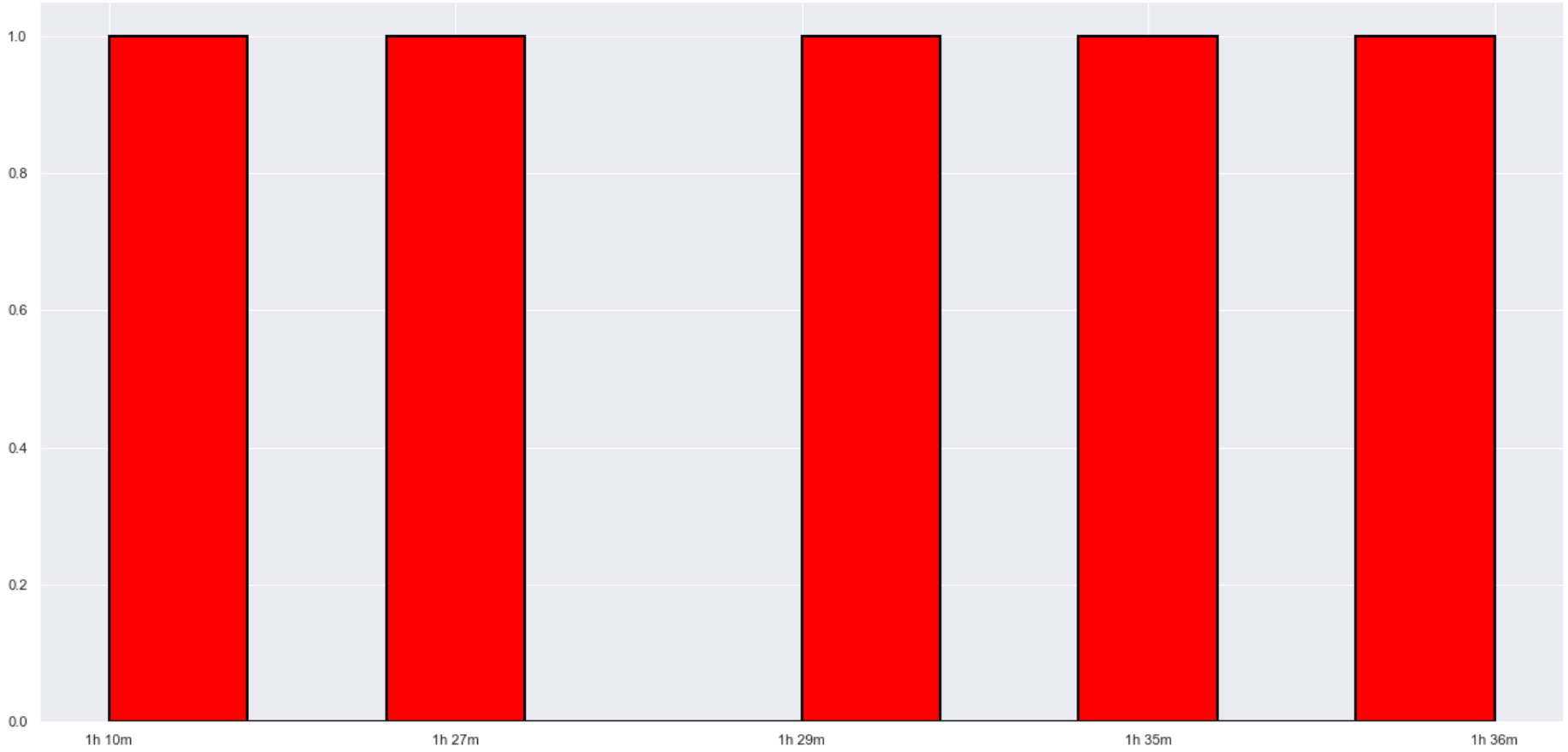
In [45]:
```
s4=df4.sort_values(by='Runtime',ascending=True).head()
s4
```

Out[45]:

Name	Runtime
Beautiful Sisters: Flesh Slave	1h 10m
Hotel Transylvania: Transformania	1h 27m
Pil's Adventures	1h 29m
Restless	1h 35m
No Exit	1h 36m

In [198]:
```
plt.hist(s4,color='red',edgecolor='black',linewidth=2)
```

Out[198]:
```
(array([1., 0., 1., 0., 0., 1., 0., 1., 0., 1.]),
 array([0. , 0.4, 0.8, 1.2, 1.6, 2. , 2.4, 2.8, 3.2, 3.6, 4. ]),
 <BarContainer object of 10 artists>)
```



In [47]:
```
df5=data.groupby(['Director'])[['Rating']].max()
df5
```

Out[47]:

Director	Rating
Adam Berg	62.0
Anthony Hayes	64.0
Arie Posin	71.0
Charise Castro Smith	77.0
Damien Power	67.0
Domee Shi	75.0
Genndy Tartakovsky	71.0
Graham Moore	72.0
Jared Bush	77.0
Josh Miller	77.0
Julien Fournet	68.0
Katsuhiko Fujii	52.0
Mark Williams	60.0
Matt Reeves	79.0
Maximilian Elfeldt	60.0
Na Hyun	62.0
Olen Steinhauer	60.0
Patrick Casey	77.0
Roland Emmerich	65.0
Régis Blondeau	60.0
Shawn Levy	70.0
Stan Lee	81.0

Director	Rating
Steve Ditko	81.0

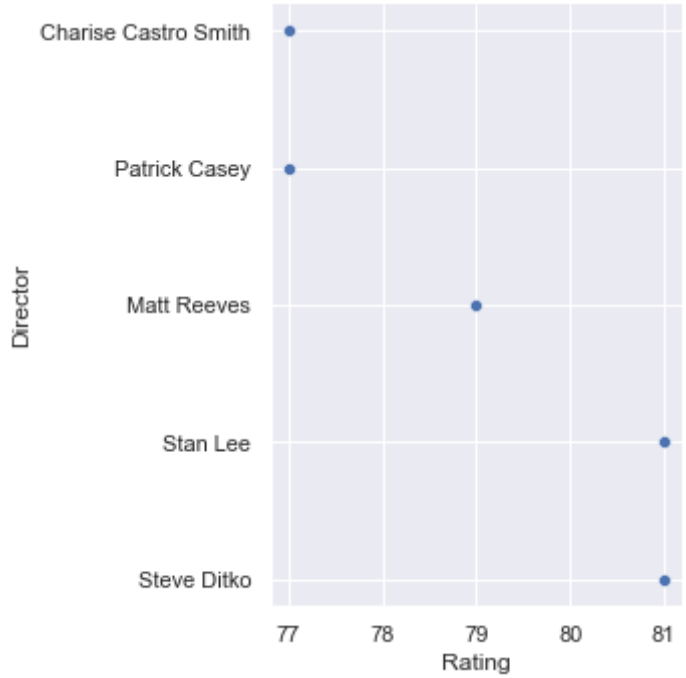
In [48]:
```
s5=df5.sort_values(by='Rating',ascending=True).tail()
s5
```

Out[48]:

Director	Rating
Charise Castro Smith	77.0
Patrick Casey	77.0
Matt Reeves	79.0
Stan Lee	81.0
Steve Ditko	81.0

In [199]:
```
plt.figure(figsize=(25,25))
sns.relplot(x='Rating',y='Director',data=s5)
plt.show()
```

```
<Figure size 1800x1800 with 0 Axes>
```



In [85]:
```
data.columns
```

Out[85]:
```
Index(['Name', 'Rating', 'Genres', 'Date', 'Runtime', 'Director', 'link'], dtype='object')
```

In [90]:
```
df6=data.groupby(['Date'])[['Genres']].max()
df6
```

Out[90]:

Date	Genres
01/13/2022	Thriller, Action
01/18/1986	Crime, Horror
02/04/2022	Action, Adventure, Science Fiction
02/11/2022	Romance, Science Fiction, Drama
02/25/2022	Horror, Thriller
03/04/2022	Crime, Mystery, Thriller
03/10/2022	Animation, Family, Comedy, Fantasy
03/11/2022	Action, Adventure, Comedy, Science Fiction
03/18/2022	Drama, Thriller, Crime
04/08/2022	Thriller, Action
08/11/2021	Adventure, Animation, Comedy, Family, Fantasy
11/24/2021	Animation, Comedy, Family, Fantasy
12/17/2021	Action, Adventure, Science Fiction
12/22/2021	Science Fiction, Action

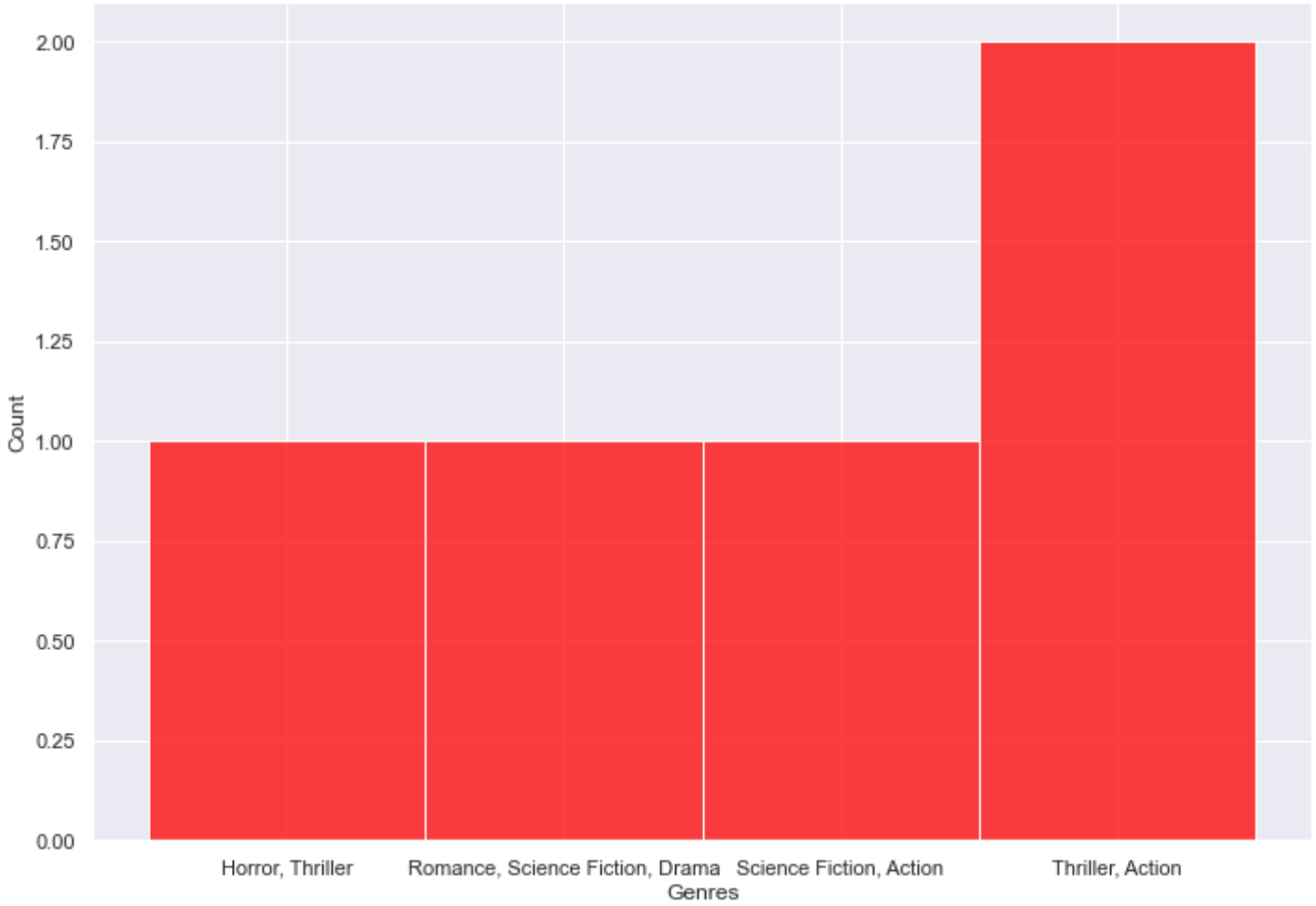
In [93]:
```
s6=df6.sort_values(by='Genres',ascending=True).tail()
s6
```

Out[93]:

Date	Genres
02/25/2022	Horror, Thriller
02/11/2022	Romance, Science Fiction, Drama
12/22/2021	Science Fiction, Action
01/13/2022	Thriller, Action
04/08/2022	Thriller, Action

In [212]:
```
sns.set(rc={'figure.figsize':(11.7,8.27)})
sns.histplot(data=s6,x='Genres',color='red')
```

Out[212]:
```
<AxesSubplot:xlabel='Genres', ylabel='Count'>
```



In [ ]:

In [138]:
```
data['Date'] = pd.to_datetime(data['Date'])
data['Date']
```

Out[138]:
```
0      2022-03-04
1      2022-03-18
2      2021-12-17
3      2022-03-10
4      2022-04-08
          ...
3995   2022-01-13
3996   2022-02-25
3997   2022-02-25
3998   1986-01-18
3999   2022-02-11
Name: Date, Length: 4000, dtype: datetime64[ns]
```

In [156]:
```
data['year'] = data['Date'].dt.year
y_data = data['year'].value_counts().to_frame()
y_data
#data['year'] = data['Date'].dt.year
#r_data = data['Rating'].value_counts().to_frame()
#r_data
```

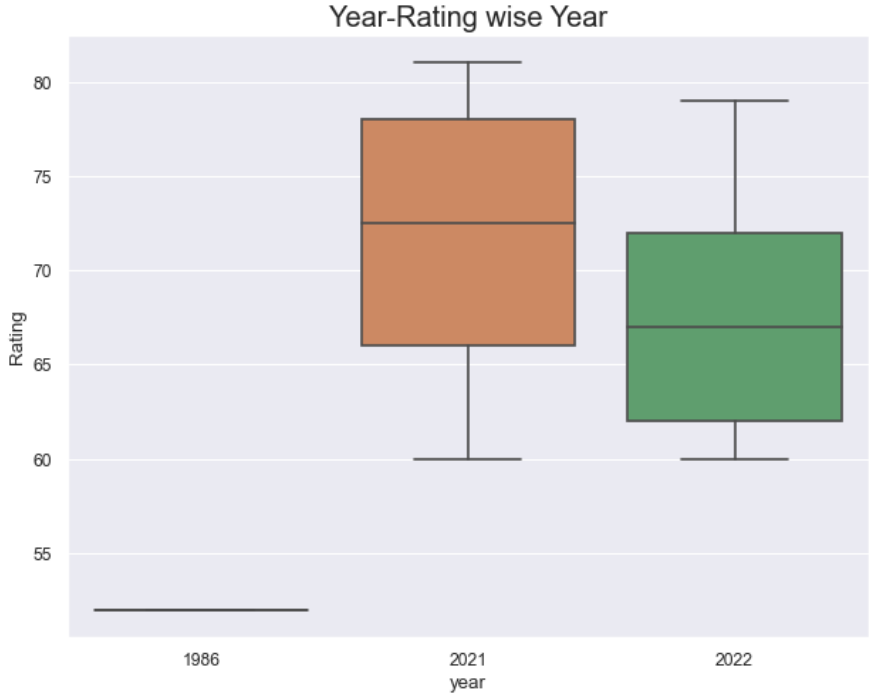
Out[156]:

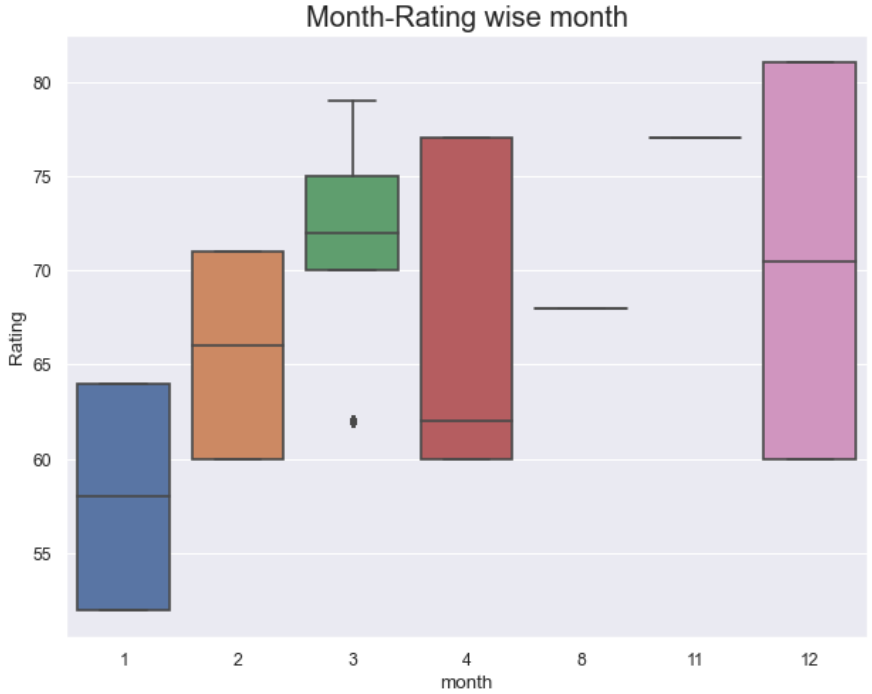
	year
2022	3000
2021	800
1986	200

In [159]:
```
data['year'] = data['Date'].dt.year
data['month'] = data['Date'].dt.month
years = data['year'].unique()


fig, axes = plt.subplots(1, 2, figsize=(20,7), dpi= 80)
sns.boxplot(x=data['year'], y=data['Rating'], ax=axes[0])
sns.boxplot(x=data['month'], y=data['Rating'], ax=axes[1])


axes[0].set_title('Year-Rating wise Year', fontsize=18);
axes[1].set_title('Month-Rating wise month', fontsize=18)
plt.show()
```





In [ ]:

In [ ]:

In [ ]: