

```
In [22]: 1 import requests
```

```
In [23]: 1 header = {'User-Agent': 'Mozilla/5.0(windows NT 6.3;Win64;x64)AppleWebKit/537.36(KHTML,like Gecko)Chrome/99.0.4844.82Safari/537.36'}
2
3 html_response = requests.get('https://www.themoviedb.org/movie',headers=header)
```

```
In [24]: 1 html_response
```

```
Out[24]: <Response [200]>
```

```
In [25]: 1 content = html_response.text
```

```
In [26]: 1 from bs4 import BeautifulSoup
```

```
In [27]: 1 soup = BeautifulSoup(content, 'lxml')
```

```
In [28]: 1 print(soup)

rel="apple-touch-icon" sizes="180x180"/>
<link href="/assets/2/favicon-32x32-543a21832c8931d3494a68881f6afcafc58e96c5d324345377f3197a37b367b5.png" rel="icon" sizes="32x32" type="image/png"/>
<link href="/assets/2/favicon-16x16-b362d267873ce9c5a39f686a11fe67fec2a72ed25fa8396c11b71aa43c938b11.png" rel="icon" sizes="16x16" type="image/png"/>
<link href="/manifest.json?version=3" rel="manifest"/>
<link href="https://www.themoviedb.org/movie" rel="canonical"/>
<link href="/opensearch.xml" rel="search" title="TMDB Search" type="application/opensearchdescription+xml"/>
<link as="font" crossorigin="" href="/assets/2/Consensus-3cba2c4d050ea63dbf7783173d288faf9ecb2942515a5e7f6e1beecabb2eaf72.woff2" rel="preload" type="font/woff2"/>
<link as="font" crossorigin="" href="/assets/2/source-sans-pro-v14-vietnamese_latin-ext_latin_greek-ext_greek_cyrillic-ext_cyrillic-regular-f38472b9e378b8c936f55a92677f575ab69e414d7cc3706ecfeacd994b7975.woff2" rel="preload" type="font/woff2"/>
<link as="font" crossorigin="" href="/assets/2/source-sans-pro-v14-vietnamese_latin-ext_latin_greek-ext_greek_cyrillic-ext_cyrillic-italic-48273a74acbc1611b6ef1d0a771bb8ef042b9922b51d5bff560571ddcb5542a.woff2" rel="preload" type="font/woff2"/>
<link as="font" crossorigin="" href="/assets/2/source-sans-pro-v14-vietnamese_latin-ext_latin_greek-ext_greek_cyrillic-ext_cyrillic-300-db9f15a13a9f8b8bc46db7756ab863e8834bd20a7a1124c4c0a5a8f7892f3e99.woff2" rel="preload" type="font/woff2"/>
<link as="font" crossorigin="" href="/assets/2/source-sans-pro-v14-vietnamese_latin-ext_latin_greek-ext_greek_cyrillic-ext_cyrillic-400-db9f15a13a9f8b8bc46db7756ab863e8834bd20a7a1124c4c0a5a8f7892f3e99.woff2" rel="preload" type="font/woff2"/>
```

```
In [ ]:
```

```
1
```

In [29]:

```
1 page_1_movie = soup.find_all('div',class_='wrapper')
2 for item in page_1_movie:
3     movie_name = soup.find('div',class_="card style_1").h2.text
4     movie_ratings = soup.find('div',class_='user_score_chart 57d1b0c49251410cae0003fd')['data-percent']
5     movie_link = soup.find('div',class_="card style_1").a['href']
6     mov_link = 'https://www.themoviedb.org/' + movie_link
7     url2 = requests.get(mov_link,headers = header)
8     mo_url = BeautifulSoup(url2.text,'lxml')
9     mo_url
10    movie_genre = mo_url.find('span',class_='genres').text.strip().replace(', ',',')
11    movie_release = mo_url.find('span',class_='release').text.strip().replace('(IN)', '')
12    movie_runtime = mo_url.find('span',class_='runtime').text.strip()
13    movie_dir = mo_url.find('li',class_='profile').a.text
14
15    my_movie_dict = {
16        'Movie Name':movie_name,
17        'Movie Ratings':movie_ratings,
18        'movie Link':mov_link,
19        'Movie Genre':movie_genre,
20        'Movie Release':movie_release,
21        'Movie Runtime':movie_runtime,
22        'Movie Dir':movie_dir
23    }
24    print(my_movie_dict)
```



```
In [40]: 1 url = 'https://www.themoviedb.org/movie?page='
```

```
In [42]: 1 lst_url = []
2 for value in range(1,201):
3     lst_url.append(url+str(value))
```

```
In [43]: 1 for val in lst_url:
```

```
2     print(val)
https://www.themoviedb.org/movie?page=148 (https://www.themoviedb.org/movie?page=148)
https://www.themoviedb.org/movie?page=149 (https://www.themoviedb.org/movie?page=149)
https://www.themoviedb.org/movie?page=150 (https://www.themoviedb.org/movie?page=150)
https://www.themoviedb.org/movie?page=151 (https://www.themoviedb.org/movie?page=151)
https://www.themoviedb.org/movie?page=152 (https://www.themoviedb.org/movie?page=152)
https://www.themoviedb.org/movie?page=153 (https://www.themoviedb.org/movie?page=153)
https://www.themoviedb.org/movie?page=154 (https://www.themoviedb.org/movie?page=154)
https://www.themoviedb.org/movie?page=155 (https://www.themoviedb.org/movie?page=155)
https://www.themoviedb.org/movie?page=156 (https://www.themoviedb.org/movie?page=156)
https://www.themoviedb.org/movie?page=157 (https://www.themoviedb.org/movie?page=157)
https://www.themoviedb.org/movie?page=158 (https://www.themoviedb.org/movie?page=158)
https://www.themoviedb.org/movie?page=159 (https://www.themoviedb.org/movie?page=159)
https://www.themoviedb.org/movie?page=160 (https://www.themoviedb.org/movie?page=160)
https://www.themoviedb.org/movie?page=161 (https://www.themoviedb.org/movie?page=161)

https://www.themoviedb.org/movie?page=162 (https://www.themoviedb.org/movie?page=162)
https://www.themoviedb.org/movie?page=163 (https://www.themoviedb.org/movie?page=163)
https://www.themoviedb.org/movie?page=164 (https://www.themoviedb.org/movie?page=164)
https://www.themoviedb.org/movie?page=165 (https://www.themoviedb.org/movie?page=165)
https://www.themoviedb.org/movie?page=166 (https://www.themoviedb.org/movie?page=166)
https://www.themoviedb.org/movie?page=167 (https://www.themoviedb.org/movie?page=167)
```

```
In [44]: 1 ldt = []
2 def almv():
3     for link in lst_url:
4         response = requests.get(link,headers= header)
5         html_text = response.text
6         movie_soup = BeautifulSoup(html_text,'lxml')
7         movie_list = movie_soup.find_all('div',class_='card style_1')
8         for item in movie_list:
9             movie_name = item.find('div',class_="content").h2.text
10            movie_ratings = item.find('div',class_='user_score_chart')['data-percent']
11            movie_link = item.find('div',class_="content").a['href']
12            mov_link = 'https://www.themoviedb.org/' + movie_link
13            url2 = requests.get(mov_link,headers = header)
14            mo_url = BeautifulSoup(url2.text,'lxml')
15            mo_url
16            movie_genre = mo_url.find('span',class_='genres').text.strip().replace(',','')
17            movie_release = mo_url.find('span',class_="release").text.strip().replace('(IN)', '')
18            movie_runtime = mo_url.find('span',class_="runtime")
19            if movie_runtime is not None:
20                movie_runtime=(movie_runtime.text.strip())
21            movie_dir = mo_url.find('li',class_='profile')
22            if movie_dir is not None:
23                movie_dir=(movie_dir.a.text)
24
25            my_movie_dict = {
26                'Movie Name':movie_name,
27                'Movie Ratings':movie_ratings,
28                'movie Link':mov_link,
29                'Movie Genre':movie_genre,
30                'Movie Release':movie_release,
31                'Movie Runtime':movie_runtime,
32                'Movie Dir':movie_dir
33            }
34            ldt.append(my_movie_dict)
35 almv()
```

```
In [45]: 1 import pandas as pd
```

```
In [46]: 1 movie_df = pd.DataFrame(ldt)
```

In [47]: 1 | movie_df

Out[47]:

	Movie Name	Movie Ratings	movie Link	Movie Genre	Movie Release	Movie Runtime	Movie Dir
0	The Batman	79.0	https://www.themoviedb.org//movie/414906	Crime, Mystery, Thriller	03/04/2022	2h 56m	Matt Reeves
1	The Outfit	70	https://www.themoviedb.org//movie/799876	Drama, Thriller, Crime	03/18/2022 (US)	1h 45m	Graham Moore
2	Spider-Man: No Way Home	81.0	https://www.themoviedb.org//movie/634649	Action, Adventure, Science Fiction	12/17/2021	2h 28m	Stan Lee
3	Yaksha: Ruthless Operations	60	https://www.themoviedb.org//movie/606402	Action	04/08/2022 (KR)	2h 5m	Na Hyun
4	Turning Red	75.0	https://www.themoviedb.org//movie/508947	Animation, Family, Comedy, Fantasy	03/10/2022 (CZ)	1h 40m	Domee Shi
...
3995	Little	68.0	https://www.themoviedb.org//movie/526050	Comedy, Fantasy, Romance	04/12/2019 (US)	1h 48m	Tina Gordon Chism
3996	The Jurassic Games	46.0	https://www.themoviedb.org//movie/504982	Science Fiction, Action, Horror	05/21/2018 (US)	1h 26m	Ryan Bellgardt
3997	Bad Luck Banging or Loony Porn	65.0	https://www.themoviedb.org//movie/790496	Comedy, Drama	11/26/2021 (RO)	1h 46m	Radu Jude
3998	Winnie the Pooh: Springtime with Roo	66.0	https://www.themoviedb.org//movie/16394	Family, Animation	03/09/2004 (US)	1h 5m	A. A. Milne
3999	Midnight in Paris	75.0	https://www.themoviedb.org//movie/59436	Fantasy, Comedy, Romance	05/20/2011 (US)	1h 34m	Woody Allen

4000 rows × 7 columns

```
In [48]: 1 pd.DataFrame(ldt)
```

Out[48]:

	Movie Name	Movie Ratings		movie Link	Movie Genre	Movie Release	Movie Runtime	Movie Dir
0	The Batman	79.0	https://www.themoviedb.org//movie/414906		Crime, Mystery, Thriller	03/04/2022	2h 56m	Matt Reeves
1	The Outfit	70	https://www.themoviedb.org//movie/799876		Drama, Thriller, Crime	03/18/2022 (US)	1h 45m	Graham Moore
2	Spider-Man: No Way Home	81.0	https://www.themoviedb.org//movie/634649		Action, Adventure, Science Fiction	12/17/2021	2h 28m	Stan Lee
3	Yaksha: Ruthless Operations	60	https://www.themoviedb.org//movie/606402		Action	04/08/2022 (KR)	2h 5m	Na Hyun
4	Turning Red	75.0	https://www.themoviedb.org//movie/508947	Animation, Family, Comedy, Fantasy		03/10/2022 (CZ)	1h 40m	Domee Shi
...
3995	Little	68.0	https://www.themoviedb.org//movie/526050		Comedy, Fantasy, Romance	04/12/2019 (US)	1h 48m	Tina Gordon Chism
3996	The Jurassic Games	46.0	https://www.themoviedb.org//movie/504982		Science Fiction, Action, Horror	05/21/2018 (US)	1h 26m	Ryan Bellgardt
3997	Bad Luck Banging or Loony Porn	65.0	https://www.themoviedb.org//movie/790496		Comedy, Drama	11/26/2021 (RO)	1h 46m	Radu Jude
3998	Winnie the Pooh: Springtime with Roo	66.0	https://www.themoviedb.org//movie/16394		Family, Animation	03/09/2004 (US)	1h 5m	A. A. Milne
3999	Midnight in Paris	75.0	https://www.themoviedb.org//movie/59436		Fantasy, Comedy, Romance	05/20/2011 (US)	1h 34m	Woody Allen

4000 rows × 7 columns

```
In [49]: 1 movie_df.to_csv('movie-list.csv')
```

CSV File to Data Visualization

```
In [100]: 1 import numpy as np  
2 import pandas as pd  
3 import plotly.express as px  
4 import seaborn as sns
```

```
In [101]: 1 data = pd.read_csv(r"C:\Users\omkar navale\movie-list.csv")
```

```
In [90]: 1 data_movies = data
```

In [92]:

```

1 data_movies.replace('#', '-', inplace=True)
2 data_movies

```

Out[92]:

	Unnamed: 0	Movie Name	Movie Ratings	movie Link	Movie Genre	Movie Release	Movie Runtime	Movie Title
0	0	The Batman	79.0	https://www.themoviedb.org//movie/414906	Crime, Mystery, Thriller	03/04/2022	2h 56m	M Reeve
1	1	The Outfit	70.0	https://www.themoviedb.org//movie/799876	Drama, Thriller, Crime	03/18/2022 (US)	1h 45m	Graham Moore
2	2	Spider-Man: No Way Home	81.0	https://www.themoviedb.org//movie/634649	Action, Adventure, Science Fiction	12/17/2021	2h 28m	Stan Lee
3	3	Yaksha: Ruthless Operations	60.0	https://www.themoviedb.org//movie/606402	Action	04/08/2022 (KR)	2h 5m	Na Hyun
4	4	Turning Red	75.0	https://www.themoviedb.org//movie/508947	Animation, Family, Comedy, Fantasy	03/10/2022 (CZ)	1h 40m	Domhnall Gleeson
...
3995	3995	Little	68.0	https://www.themoviedb.org//movie/526050	Comedy, Fantasy, Romance	04/12/2019 (US)	1h 48m	Ti West, Gord Chi
3996	3996	The Jurassic Games	46.0	https://www.themoviedb.org//movie/504982	Science Fiction, Action, Horror	05/21/2018 (US)	1h 26m	Ry Cuming, Bellga
3997	3997	Bad Luck Banging or Loony Porn	65.0	https://www.themoviedb.org//movie/790496	Comedy, Drama	11/26/2021 (RO)	1h 46m	Rajat Kapoor, Juhi Chawla
3998	3998	Winnie the Pooh: Springtime with Roo	66.0	https://www.themoviedb.org//movie/16394	Family, Animation	03/09/2004 (US)	1h 5m	A. Milne, E. Shepard
3999	3999	Midnight in Paris	75.0	https://www.themoviedb.org//movie/59436	Fantasy, Comedy, Romance	05/20/2011 (US)	1h 34m	Woody Allen

4000 rows × 8 columns

```
In [93]: 1 data_movies.fillna(value=0,inplace=True)
```

```
In [94]: 1 data_movies.isnull().sum()
```

```
Out[94]: Unnamed: 0      0  
Movie Name        0  
Movie Ratings     0  
movie Link        0  
Movie Genre       0  
Movie Release     0  
Movie Runtime     0  
Movie Dir         0  
dtype: int64
```

```
In [95]: 1 data_movies.describe()
```

```
Out[95]:
```

	Unnamed: 0	Movie Ratings
count	4000.000000	4000.000000
mean	1999.500000	65.452000
std	1154.844867	11.002634
min	0.000000	0.000000
25%	999.750000	60.000000
50%	1999.500000	66.000000
75%	2999.250000	72.000000
max	3999.000000	91.000000

```
In [96]: 1 data_movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4000 entries, 0 to 3999
Data columns (total 8 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Unnamed: 0        4000 non-null    int64  
 1   Movie Name       4000 non-null    object  
 2   Movie Ratings    4000 non-null    float64 
 3   movie Link       4000 non-null    object  
 4   Movie Genre      4000 non-null    object  
 5   Movie Release    4000 non-null    object  
 6   Movie Runtime    4000 non-null    object  
 7   Movie Dir        4000 non-null    object  
dtypes: float64(1), int64(1), object(6)
memory usage: 250.1+ KB
```

```
In [77]: 1 data.columns
```

```
Out[77]: Index(['Unnamed: 0', 'Movie Name', 'Movie Ratings', 'movie Link',
                 'Movie Genre', 'Movie Release', 'Movie Runtime', 'Movie Dir'],
                dtype='object')
```

In [78]: 1 data_movies.head(5)

Out[78]:

	Unnamed: 0	Movie Name	Movie Ratings	movie Link	Movie Genre	Movie Release	Movie Runtime	Movie Dir
0	0	The Batman	79.0	https://www.themoviedb.org//movie/414906	Crime, Mystery, Thriller	03/04/2022	2h 56m	Matt Reeves
1	1	The Outfit	70.0	https://www.themoviedb.org//movie/799876	Drama, Thriller, Crime	03/18/2022 (US)	1h 45m	Graham Moore
2	2	Spider-Man: No Way Home	81.0	https://www.themoviedb.org//movie/634649	Action, Adventure, Science Fiction	12/17/2021	2h 28m	Stan Lee
3	3	Yaksha: Ruthless Operations	60.0	https://www.themoviedb.org//movie/606402	Action	04/08/2022 (KR)	2h 5m	Na Hyun
4	4	Turning Red	75.0	https://www.themoviedb.org//movie/508947	Animation, Family, Comedy, Fantasy	03/10/2022 (CZ)	1h 40m	Domee Shi

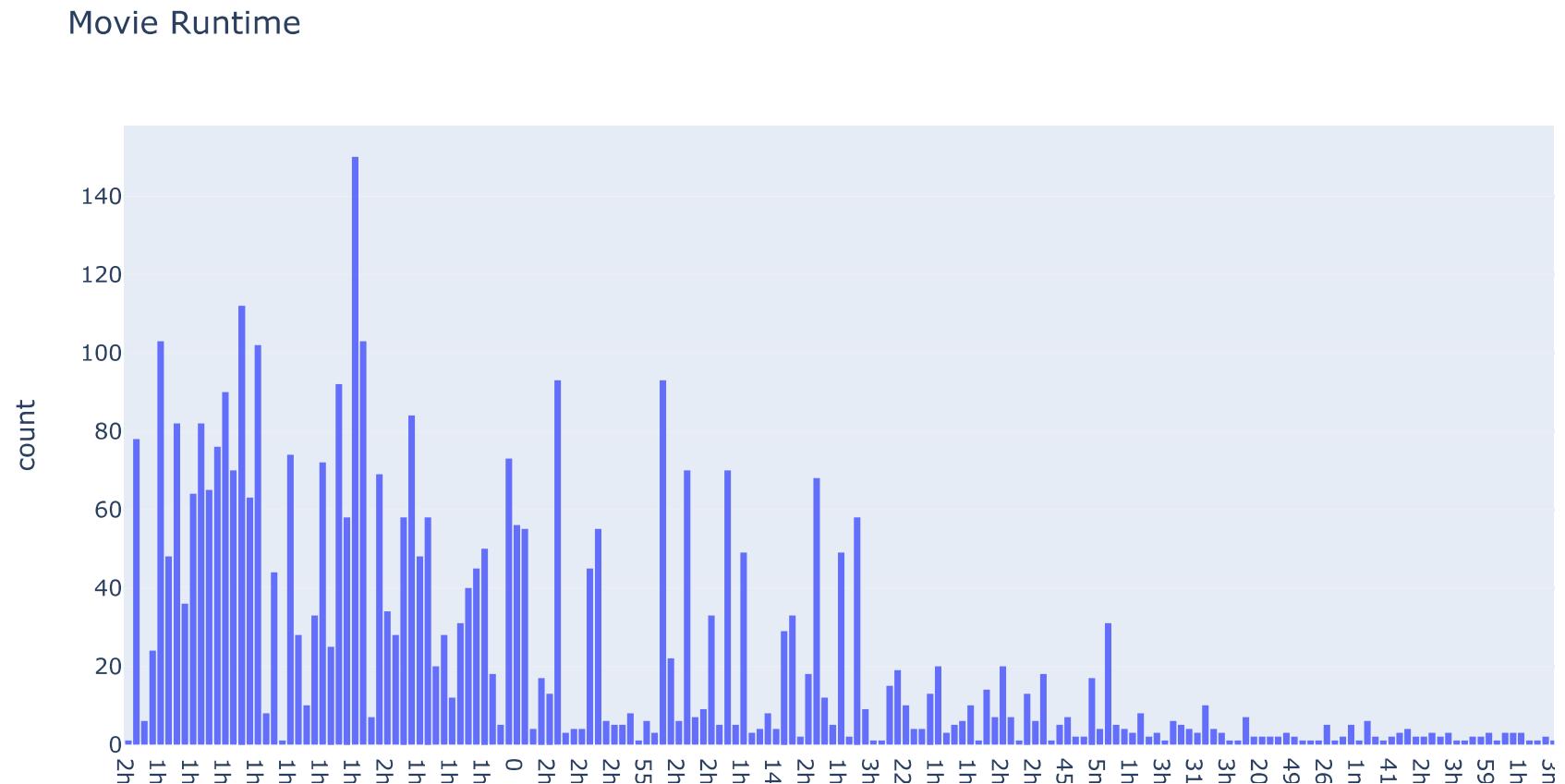
In [97]: 1 data_movies.tail(5)

Out[97]:

	Unnamed: 0	Movie Name	Movie Ratings	movie Link	Movie Genre	Movie Release	Movie Runtime	Movie Dir
3995	3995	Little	68.0	https://www.themoviedb.org//movie/526050	Comedy, Fantasy, Romance	04/12/2019 (US)	1h 48m	Tina Gordon Chism
3996	3996	The Jurassic Games	46.0	https://www.themoviedb.org//movie/504982	Science Fiction, Action, Horror	05/21/2018 (US)	1h 26m	Ryan Bellgardt
3997	3997	Bad Luck Banging or Loony Porn	65.0	https://www.themoviedb.org//movie/790496	Comedy, Drama	11/26/2021 (RO)	1h 46m	Radu Jude
3998	3998	Winnie the Pooh: Springtime with Roo	66.0	https://www.themoviedb.org//movie/16394	Family, Animation	03/09/2004 (US)	1h 5m	A. A. Milne
3999	3999	Midnight in Paris	75.0	https://www.themoviedb.org//movie/59436	Fantasy, Comedy, Romance	05/20/2011 (US)	1h 34m	Woody Allen

Histogram

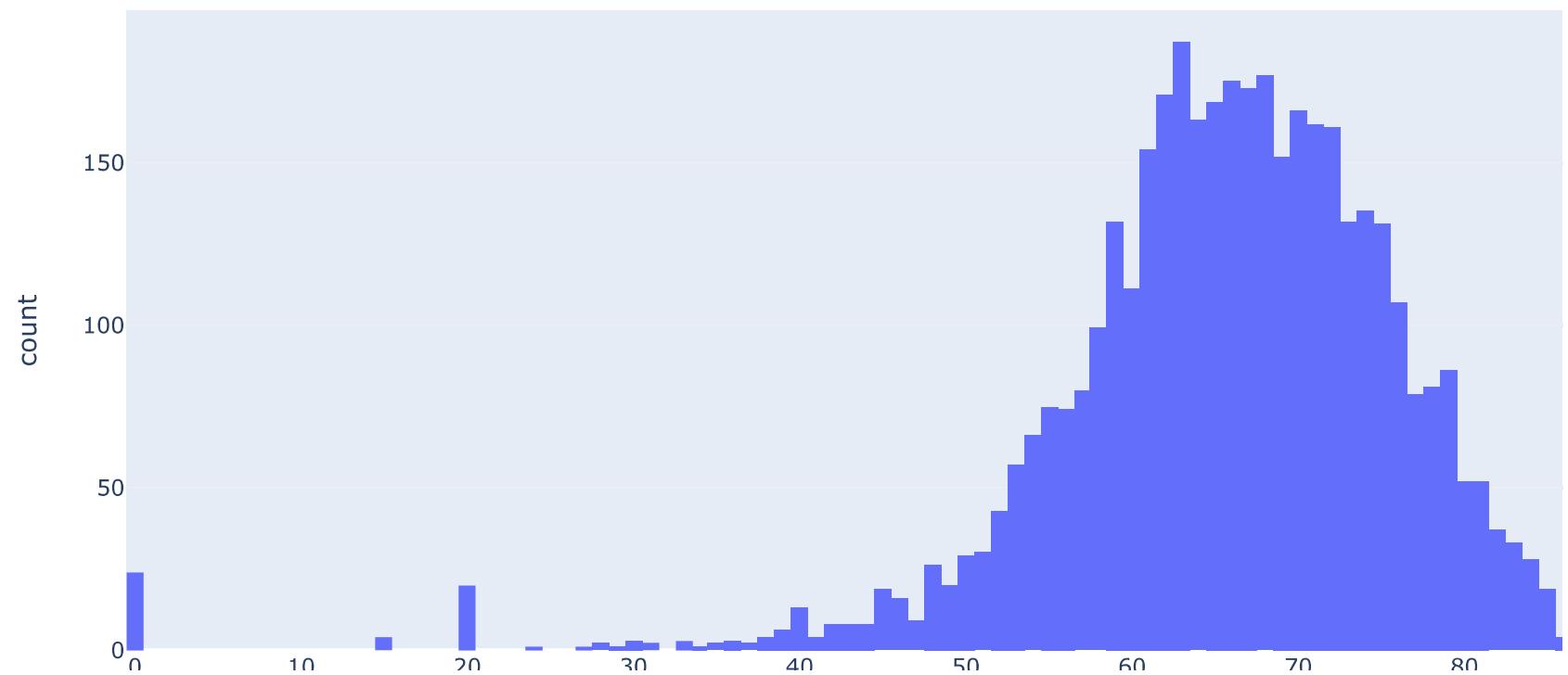
```
In [119]: 1 # Histogram in Movie Runtime  
2 px.histogram(x='Movie Runtime',title='Movie Runtime',data_frame=data_movies)
```



In [121]:

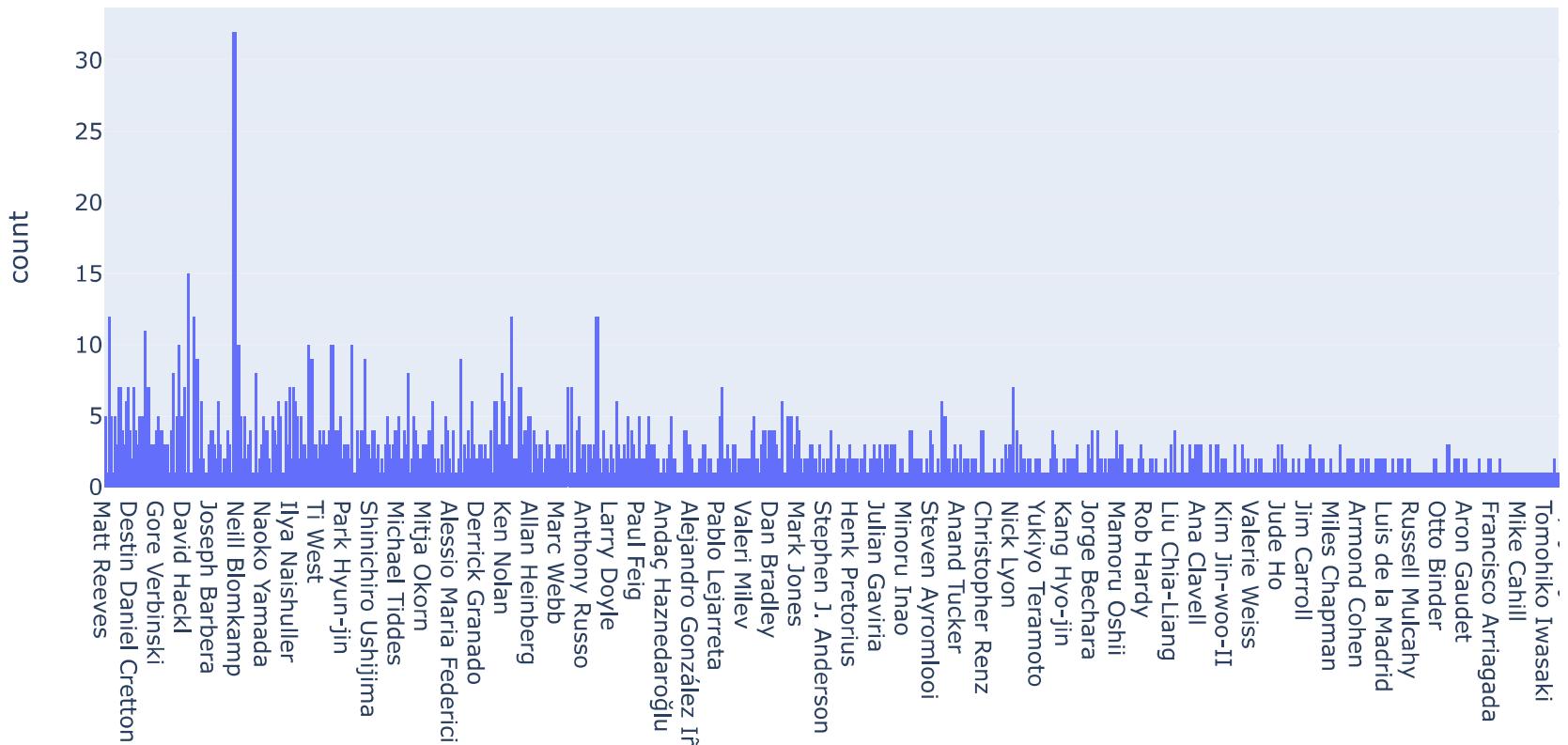
```
1 # Histogram in Movie Ratings
2 px.histogram(x='Movie Ratings',title = 'Movie Ratings',data_frame=data_movies)
```

Movie Ratings



In [123]:

```
1 #Histogram in Movie Dir
2 px.histogram(x='Movie Dir',data_frame=data_movies)
```



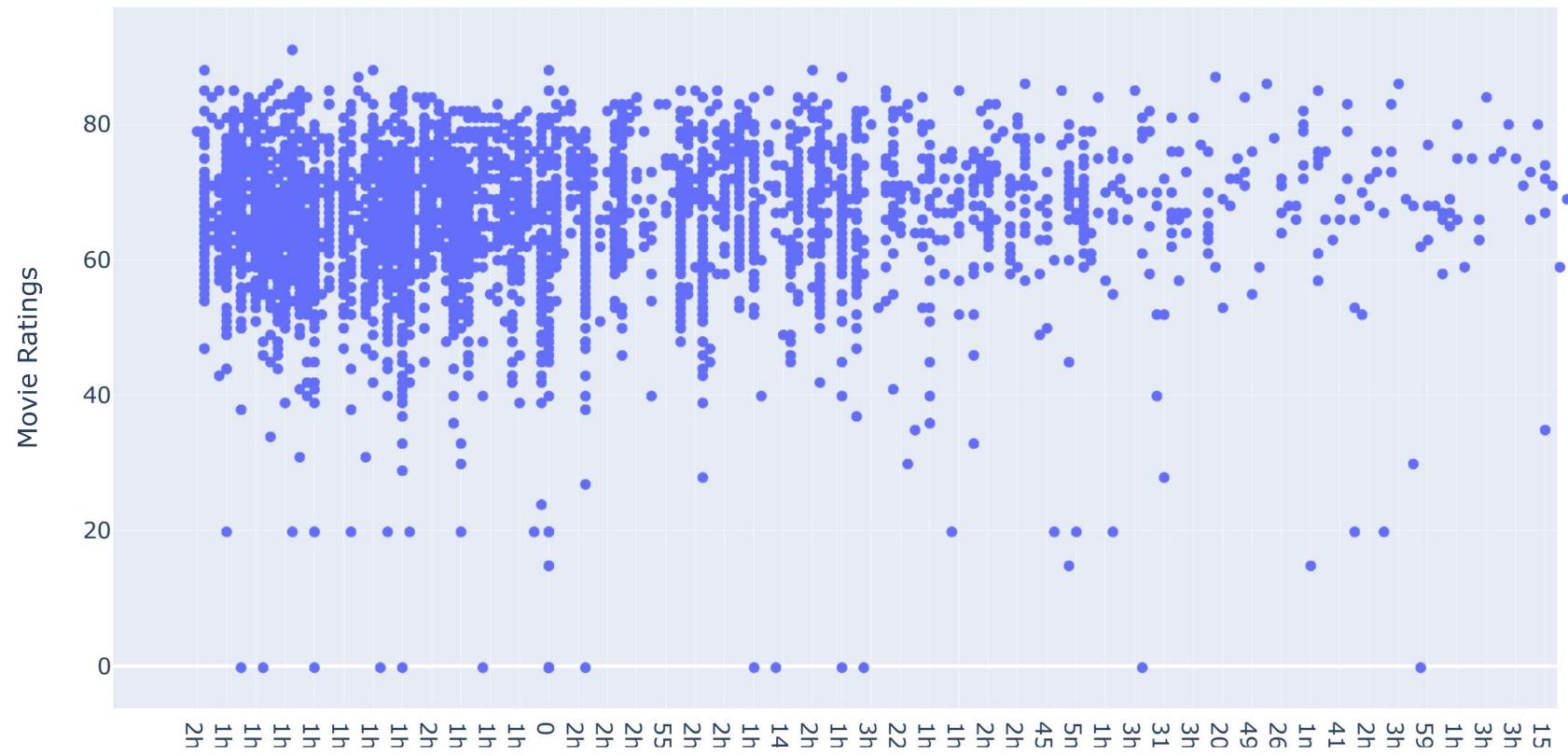
```
In [128]: 1 px.histogram(x='Movie Genre', data_frame=data_movies)
```



Scatter Plot

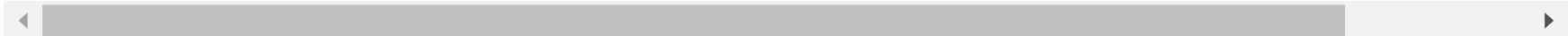
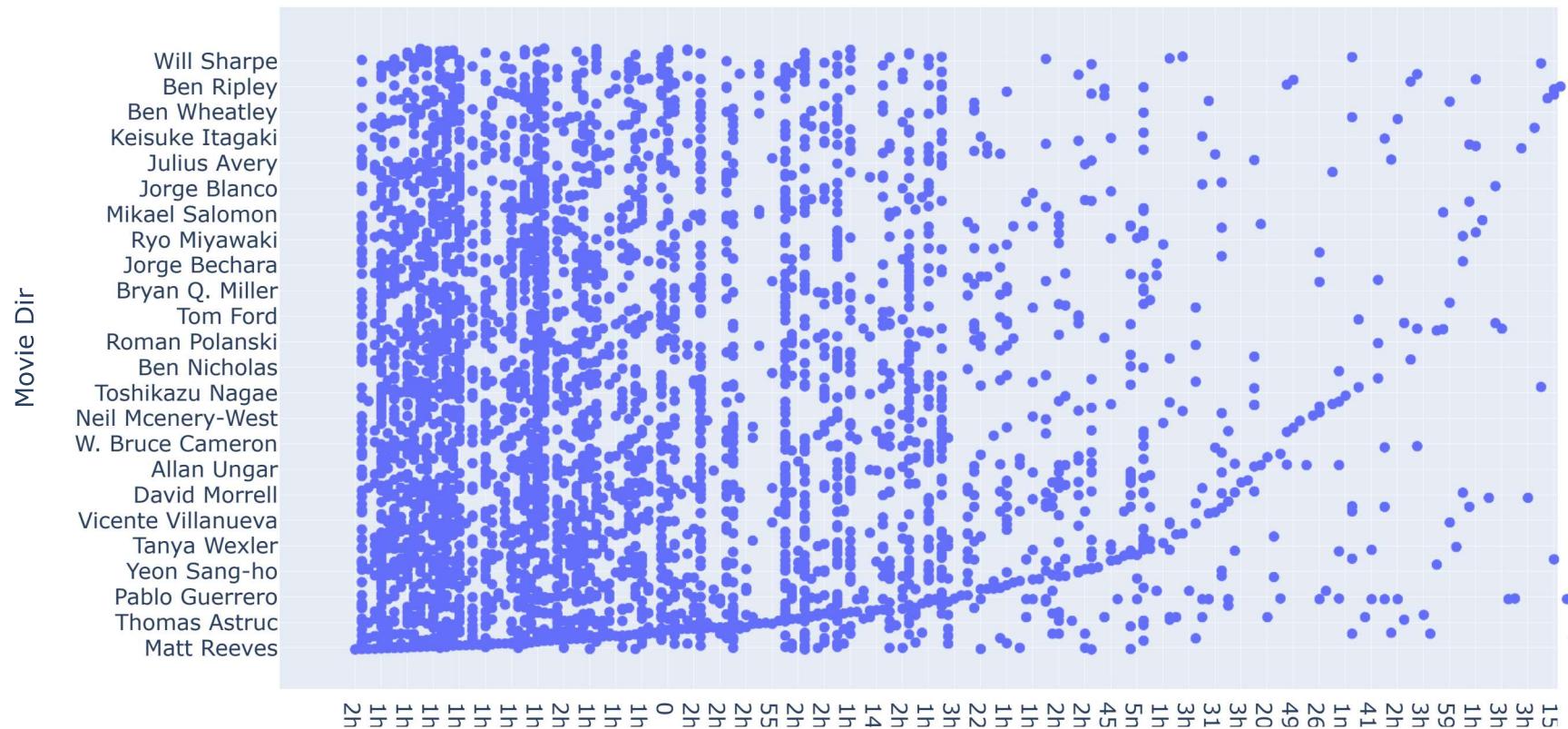
In [127]:

```
1 # Scatter plot  
2 px.scatter(x='Movie Runtime',y='Movie Ratings',data_frame=data_movies)
```

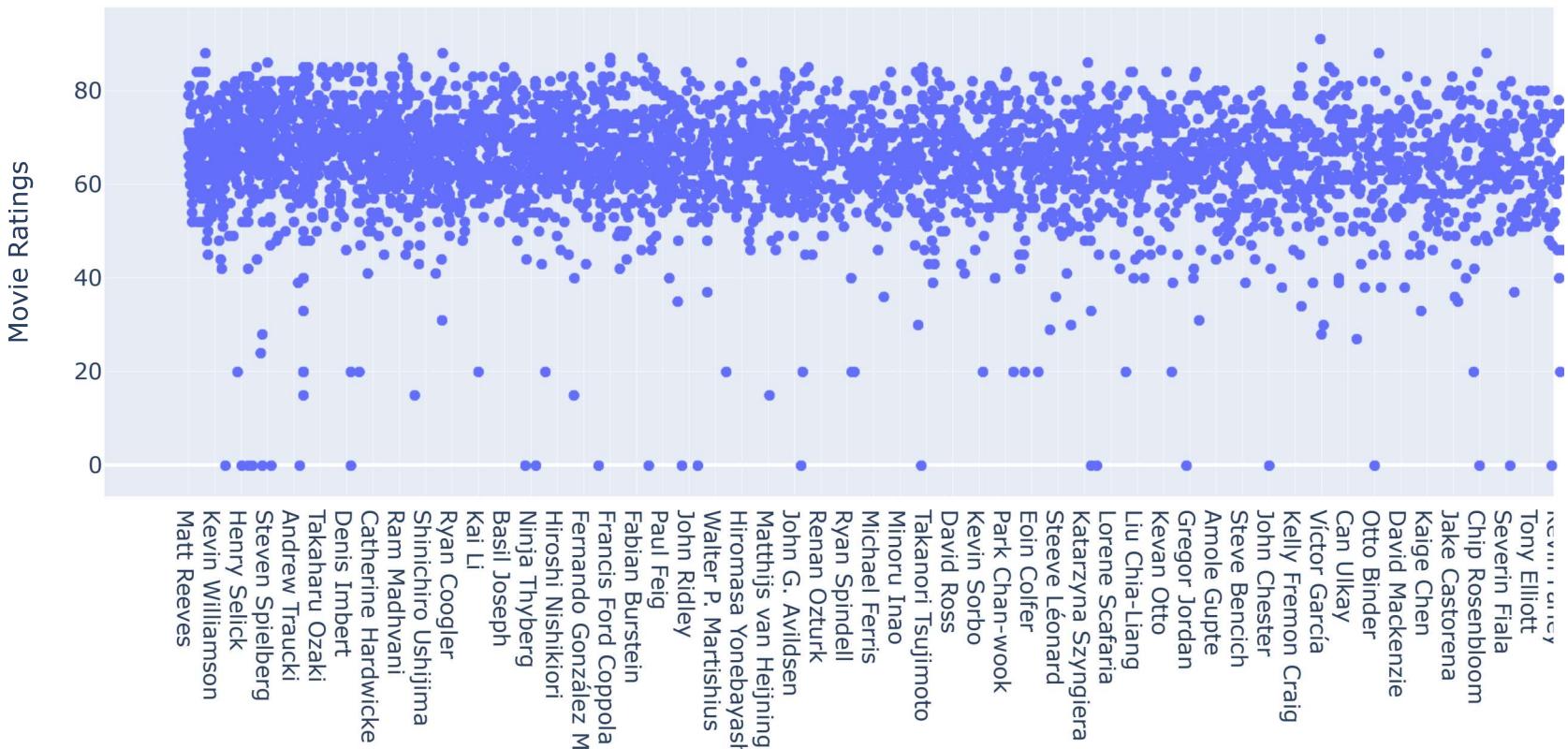


In [131]:

```
1 # Scatter Plot
2 px.scatter(x='Movie Runtime',y='Movie Dir',data_frame=data_movies)
```



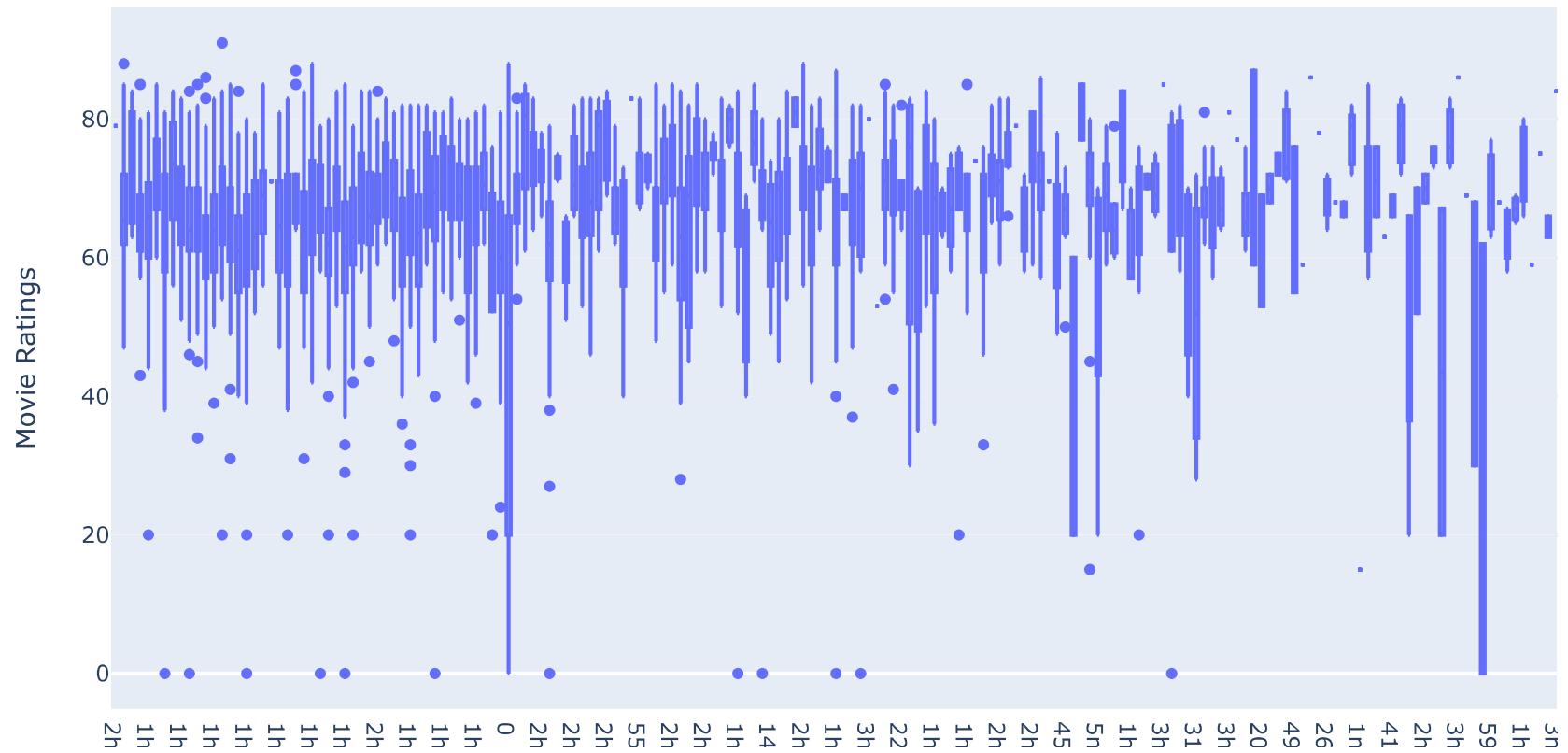
```
In [133]: 1 px.scatter(x='Movie Dir',y='Movie Ratings',data_frame=data_movies)
```



Box Plot

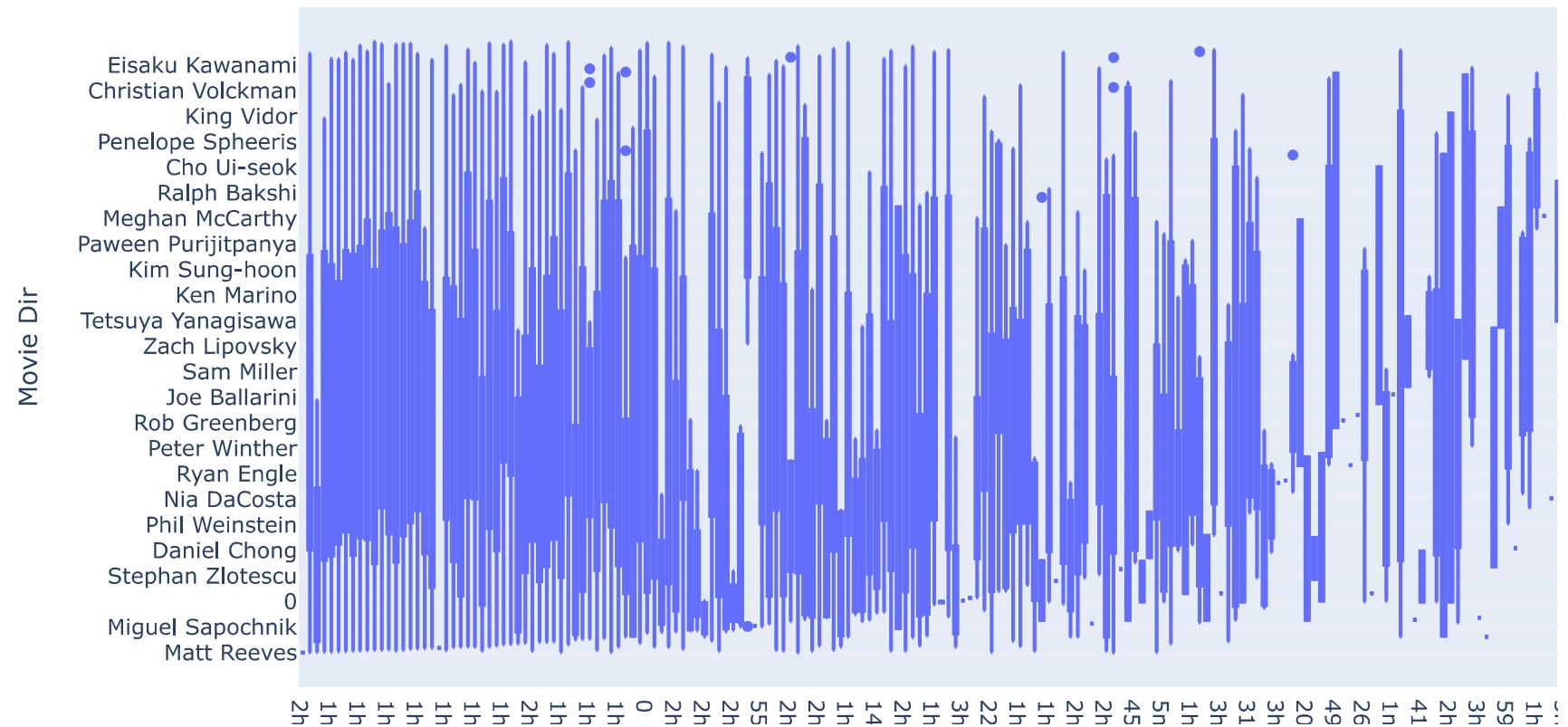
In [139]:

```
1 # Box Plot  
2 px.box(x='Movie Runtime',y='Movie Ratings',data_frame =data_movies)
```



In [140]:

```
1 # Box Plot
2 px.box(x='Movie Runtime',y='Movie Dir',data_frame =data_movies)
```



In []:

1

In []:

1

