

movie-recommendation

July 15, 2023

```
[14]: #importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('movie1.csv')
df.head(10)
```

```
[14]:      adult      belongs_to_collection      budget \
0  False  {'id': 10194, 'name': 'Toy Story Collection', ...  30000000
1  False                                     NaN  65000000
2  False  {'id': 119050, 'name': 'Grumpy Old Men Collect...    0
3  False                                     NaN  16000000
4  False  {'id': 96871, 'name': 'Father of the Bride Col...    0
5  False                                     NaN  60000000
6  False                                     NaN  58000000
7  False                                     NaN    0
8  False                                     NaN  35000000
9  False  {'id': 645, 'name': 'James Bond Collection', '...  58000000

      genres \
0  [{'id': 16, 'name': 'Animation'}, {'id': 35, 'name': 'Comedy'}]
1  [{'id': 12, 'name': 'Adventure'}, {'id': 14, 'name': 'Comedy'}]
2  [{'id': 10749, 'name': 'Romance'}, {'id': 35, 'name': 'Comedy'}]
3  [{'id': 35, 'name': 'Comedy'}, {'id': 18, 'name': 'Drama'}]
4  [{'id': 35, 'name': 'Comedy'}]
5  [{'id': 28, 'name': 'Action'}, {'id': 80, 'name': 'Mystery'}]
6  [{'id': 35, 'name': 'Comedy'}, {'id': 10749, 'name': 'Romance'}]
7  [{'id': 28, 'name': 'Action'}, {'id': 12, 'name': 'Adventure'}]
8  [{'id': 28, 'name': 'Action'}, {'id': 12, 'name': 'Adventure'}]
9  [{'id': 12, 'name': 'Adventure'}, {'id': 28, 'name': 'Action'}]

      homepage      id      imdb_id \
0  http://toystory.disney.com/toy-story    862  tt0114709
1                                     NaN    8844  tt0113497
2                                     NaN    15602  tt0113228
```

3		NaN	31357	tt0114885
4		NaN	11862	tt0113041
5		NaN	949	tt0113277
6		NaN	11860	tt0114319
7		NaN	45325	tt0112302
8		NaN	9091	tt0114576
9	http://www.mgm.com/view/movie/757/Goldeneye/		710	tt0113189

	original_language	original_title	\
0	en	Toy Story	
1	en	Jumanji	
2	en	Grumpier Old Men	
3	en	Waiting to Exhale	
4	en	Father of the Bride Part II	
5	en	Heat	
6	en	Sabrina	
7	en	Tom and Huck	
8	en	Sudden Death	
9	en	GoldenEye	

	overview	...	release_date	\
0	Led by Woody, Andy's toys live happily in his	10/30/1995	
1	When siblings Judy and Peter discover an encha...	...	12/15/1995	
2	A family wedding reignites the ancient feud be...	...	12/22/1995	
3	Cheated on, mistreated and stepped on, the wom...	...	12/22/1995	
4	Just when George Banks has recovered from his	2/10/1995	
5	Obsessive master thief, Neil McCauley leads a	12/15/1995	
6	An ugly duckling having undergone a remarkable...	...	12/15/1995	
7	A mischievous young boy, Tom Sawyer, witnesses...	...	12/22/1995	
8	International action superstar Jean Claude Van...	...	12/22/1995	
9	James Bond must unmask the mysterious head of	11/16/1995	

	revenue	runtime	spoken_languages	\
0	373554033	81.0	[{'iso_639_1': 'en', 'name': 'English'}]	
1	262797249	104.0	[{'iso_639_1': 'en', 'name': 'English'}, {'iso...	
2	0	101.0	[{'iso_639_1': 'en', 'name': 'English'}]	
3	81452156	127.0	[{'iso_639_1': 'en', 'name': 'English'}]	
4	76578911	106.0	[{'iso_639_1': 'en', 'name': 'English'}]	
5	187436818	170.0	[{'iso_639_1': 'en', 'name': 'English'}, {'iso...	
6	0	127.0	[{'iso_639_1': 'fr', 'name': 'Français'}, {'is...	
7	0	97.0	[{'iso_639_1': 'en', 'name': 'English'}, {'iso...	
8	64350171	106.0	[{'iso_639_1': 'en', 'name': 'English'}]	
9	352194034	130.0	[{'iso_639_1': 'en', 'name': 'English'}, {'iso...	

	status	tagline	\
0	Released	NaN	
1	Released	Roll the dice and unleash the excitement!	

```

2 Released Still Yelling. Still Fighting. Still Ready for...
3 Released Friends are the people who let you be yourself...
4 Released Just When His World Is Back To Normal... He's ...
5 Released A Los Angeles Crime Saga
6 Released You are cordially invited to the most surprisi...
7 Released The Original Bad Boys.
8 Released Terror goes into overtime.
9 Released No limits. No fears. No substitutes.

```

	title	video	vote_average	vote_count
0	Toy Story	False	7.7	5415
1	Jumanji	False	6.9	2413
2	Grumpier Old Men	False	6.5	92
3	Waiting to Exhale	False	6.1	34
4	Father of the Bride Part II	False	5.7	173
5	Heat	False	7.7	1886
6	Sabrina	False	6.2	141
7	Tom and Huck	False	5.4	45
8	Sudden Death	False	5.5	174
9	GoldenEye	False	6.6	1194

[10 rows x 24 columns]

```
[15]: df.tail(10)
```

```

[15]:      adult belongs_to_collection  budget  \

1100  False                      NaN  400000
1101  False                      NaN     0
1102  False                      NaN     0
1103  False                      NaN     0
1104  False                      NaN     0
1105  False                      NaN  500000
1106  False                      NaN     0
1107  False                      NaN     0
1108  False                      NaN     0
1109  False                      NaN     0

      genres  \

1100  [{'id': 12, 'name': 'Adventure'}, {'id': 35, 'name': 'Comedy'}, {'id': 18, 'name': 'Drama'}, {'id': 10749, 'name': 'Romance'}, {'id': 18, 'name': 'Drama'}, {'id': 99, 'name': 'Documentary'}]
1101  [{'id': 10749, 'name': 'Romance'}, {'id': 18, 'name': 'Drama'}, {'id': 99, 'name': 'Documentary'}]
1102  []
1103  [{'id': 18, 'name': 'Drama'}]
1104  [{'id': 18, 'name': 'Drama'}]
1105  [{'id': 35, 'name': 'Comedy'}, {'id': 80, 'name': 'Drama'}, {'id': 99, 'name': 'Documentary'}]
1106  [{'id': 99, 'name': 'Documentary'}]
1107  [{'id': 99, 'name': 'Documentary'}]
1108  []

```

1109 [{"id": 99, 'name': 'Documentary'}]

	homepage	id	imdb_id \
1100	http://www.sonypictures.com/cthe/montypython/	762	tt0071853
1101	NaN	44497	tt0116587
1102	NaN	85588	tt0112777
1103	NaN	278621	tt0116245
1104	NaN	99479	tt0116212
1105	NaN	55049	tt0117968
1106	NaN	124829	tt0116886
1107	NaN	79306	tt0117677
1108	NaN	356054	tt0112759
1109	NaN	10548	tt0118147

	original_language	original_title \
1100	en	Monty Python and the Holy Grail
1101	en	Hustler White
1102	en	Dadetown
1103	en	Everything Relative
1104	en	Entertaining Angels - The Dorothy Day Story
1105	en	Trois vies et une seule mort
1106	en	The Line King: The Al Hirschfeld Story
1107	de	Snowriders
1108	en	Curtis's Charm
1109	en	When We Were Kings

	overview ...	release_date \
1100	King Arthur, accompanied by his squire, recrui...	3/13/1975
1101	Reminiscent of Sunset Boulevard, Hustler White...	7/19/1996
1102	Interviews in the Michael Moore/"Roger and Me"...	9/2/1995
1103	Old college chums get together for a weekend r...	9/13/1996
1104	"Entertaining Angels" is an interesting title ...	9/27/1996
1105	Raoul Ruiz's enthralling thriller is made up o...	5/15/1996
1106	The childhood, adolescence, and incredible adu...	9/27/1996
1107	Spectacular shots of snowcapped peaks and extr...	10/17/1996
1108	This film, based on a Jim Carroll short story,...	9/18/1995
1109	It's 1974. Muhammad Ali is 32 and thought by m...	10/25/1996

	revenue	runtime	spoken_languages \
1100	5028948	91.0	[{'iso_639_1': 'fr', 'name': 'Français'}, {'is...
1101	127251	79.0	[{'iso_639_1': 'en', 'name': 'English'}]
1102	0	93.0	[]
1103	0	110.0	[{'iso_639_1': 'en', 'name': 'English'}]
1104	0	112.0	[]
1105	0	123.0	[{'iso_639_1': 'fr', 'name': 'Français'}]
1106	0	86.0	[]
1107	2035470	96.0	[{'iso_639_1': 'en', 'name': 'English'}]

```

1108      0      74.0      []
1109 2666118    89.0      [{'iso_639_1': 'en', 'name': 'English'}]

```

```

      status      tagline \
1100 Released And now! At Last! Another film completely diff...
1101 Released      See it with someone you've paid for.
1102 Released      NaN
1103 Released      NaN
1104 Released      NaN
1105 Released      NaN
1106 Released      NaN
1107 Released      Man Wasn't Meant To Fly... Until Now!
1108 Released      NaN
1109 Released      The untold story of the Rumble in the Jungle.

```

```

      title  video  vote_average \
1100      Monty Python and the Holy Grail  False      7.8
1101      Hustler White  False      6.8
1102      Dadetown  False      8.0
1103      Everything Relative  False      4.0
1104 Entertaining Angels - The Dorothy Day Story  False      5.0
1105      Three Lives and Only One Death  False      8.0
1106      The Line King: The Al Hirschfeld Story  False      0.0
1107      Snowriders  False      0.0
1108      Curtis's Charm  False      0.0
1109      When We Were Kings  False      7.3

```

```

      vote_count
1100      1742
1101      12
1102      1
1103      1
1104      1
1105      5
1106      0
1107      0
1108      0
1109      54

```

```
[10 rows x 24 columns]
```

```
[62]: import seaborn as sns
sns.pairplot(df)
```

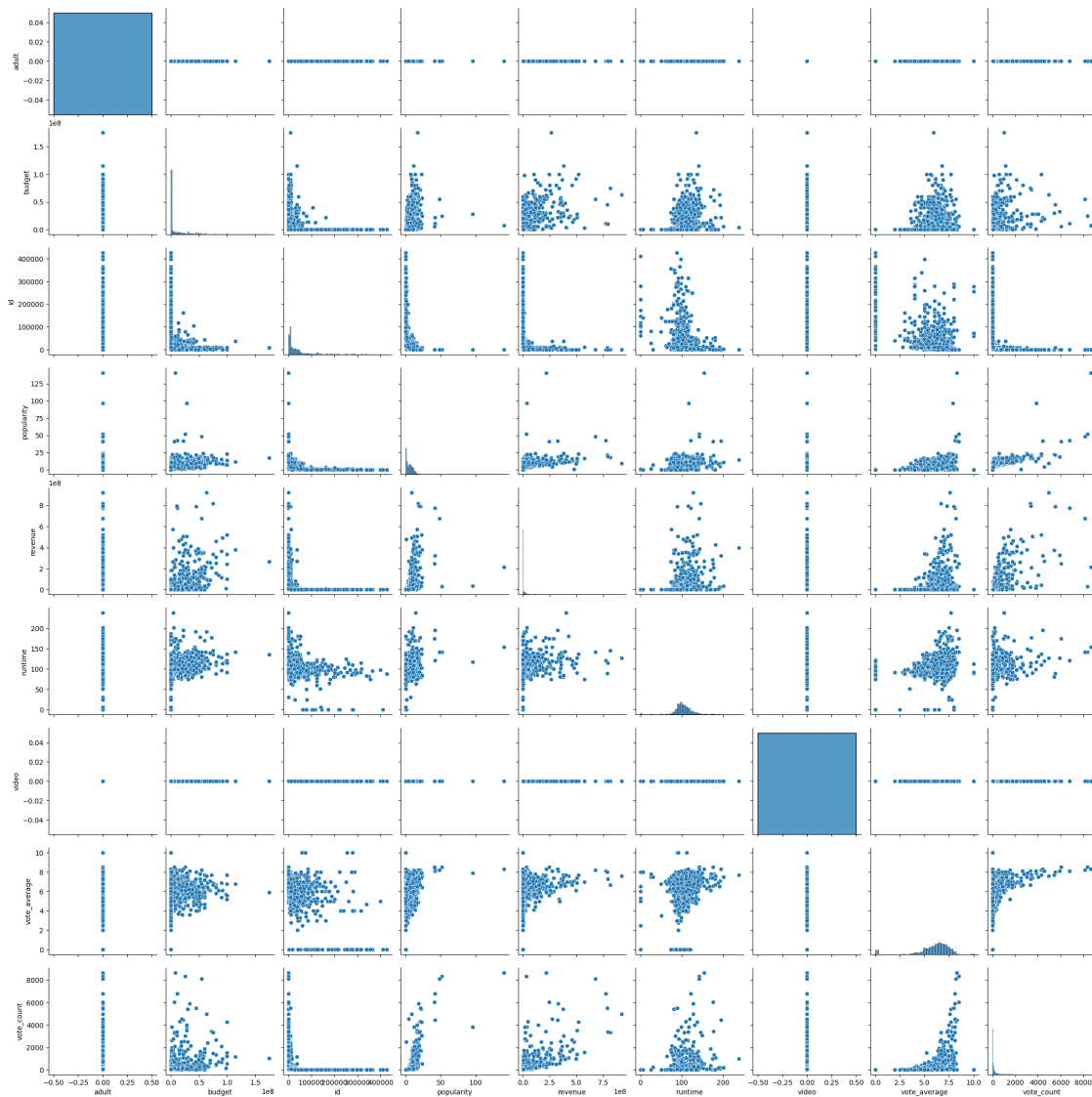
```

<__array_function__ internals>:180: RuntimeWarning: Converting input from bool
to <class 'numpy.uint8'> for compatibility.
<__array_function__ internals>:180: RuntimeWarning: Converting input from bool

```

to `<class 'numpy.uint8'>` for compatibility.

[62]: `<seaborn.axisgrid.PairGrid at 0x7b565d53fbb0>`



[17]: `df.shape`

[17]: (1110, 24)

[63]: `df.describe()`

[63]:

	budget	id	popularity	revenue	runtime \
count	1.110000e+03	1110.000000	1110.000000	1.110000e+03	1105.000000
mean	9.906667e+06	38240.970270	6.125317	3.478202e+07	104.116742

std	1.894356e+07	59656.633907	7.375777	9.712548e+07	22.445303
min	0.000000e+00	5.000000	0.000082	0.000000e+00	0.000000
25%	0.000000e+00	9099.250000	1.317025	0.000000e+00	93.000000
50%	0.000000e+00	15308.000000	5.043478	0.000000e+00	102.000000
75%	1.200000e+07	41589.500000	9.170414	1.755754e+07	114.000000
max	1.750000e+08	427043.000000	140.950236	9.201000e+08	238.000000

	vote_average	vote_count
count	1110.000000	1110.000000
mean	5.994505	295.816216
std	1.648361	809.648173
min	0.000000	0.000000
25%	5.400000	10.000000
50%	6.300000	49.500000
75%	7.000000	199.000000
max	10.000000	8670.000000

```
[24]: # finding the average revenue
a = df
a = a.groupby('title')['revenue'].mean()
a
```

```
[24]: title
1-900                                0.0
2 Days in the Valley                 0.0
20,000 Leagues Under the Sea       28200000.0
2001: A Space Odyssey              68700000.0
301, 302                             0.0
...
Wolf                                0.0
Wyatt Earp                        25052000.0
Yankee Zulu                        0.0
Yes, Madam                         0.0
À la mode                          0.0
Name: revenue, Length: 1108, dtype: float64
```

```
[30]: # avg budget
a = df
a = a.groupby('title')['budget'].mean()
a
```

```
[30]: title
1-900                                0.0
2 Days in the Valley                 0.0
20,000 Leagues Under the Sea        5000000.0
2001: A Space Odyssey              10500000.0
301, 302                             0.0
```

```

...
Wolf                                0.0
Wyatt Earp                        63000000.0
Yankee Zulu                        0.0
Yes, Madam                        0.0
À la mode                         0.0
Name: budget, Length: 1108, dtype: float64

```

```

[31]: # avg popularity
a = df
a = a.groupby('title')['popularity'].mean()
a

```

```

[31]: title
1-900                                0.062128
2 Days in the Valley                4.989276
20,000 Leagues Under the Sea       8.722406
2001: A Space Odyssey              22.494622
301, 302                           0.396626

...
Wolf                               14.063845
Wyatt Earp                         9.277147
Yankee Zulu                        1.558066
Yes, Madam                        1.060097
À la mode                         0.000082
Name: popularity, Length: 1108, dtype: float64

```

```

[32]: # avg runtime
a = df
a = a.groupby('title')['runtime'].mean()
a

```

```

[32]: title
1-900                                0.0
2 Days in the Valley                104.0
20,000 Leagues Under the Sea       127.0
2001: A Space Odyssey              149.0
301, 302                           100.0

...
Wolf                               125.0
Wyatt Earp                         191.0
Yankee Zulu                        100.0
Yes, Madam                         93.0
À la mode                         82.0
Name: runtime, Length: 1108, dtype: float64

```



```
[33]: # avg count
a = df
a = a.groupby('title')['revenue'].count()
a
```

```
[33]: title
1-900 1
2 Days in the Valley 1
20,000 Leagues Under the Sea 1
2001: A Space Odyssey 1
301, 302 1
..
Wolf 1
Wyatt Earp 1
Yankee Zulu 1
Yes, Madam 1
À la mode 1
Name: revenue, Length: 1108, dtype: int64
```

```
[34]: # avg budget
a = df
a = a.groupby('title')['budget'].count()
a
```

```
[34]: title
1-900 1
2 Days in the Valley 1
20,000 Leagues Under the Sea 1
2001: A Space Odyssey 1
301, 302 1
..
Wolf 1
Wyatt Earp 1
Yankee Zulu 1
Yes, Madam 1
À la mode 1
Name: budget, Length: 1108, dtype: int64
```

```
[35]: # avg popularity
a = df
a = a.groupby('title')['popularity'].count()
a
```

```
[35]: title
1-900 1
2 Days in the Valley 1
20,000 Leagues Under the Sea 1
```

```

2001: A Space Odyssey      1
301, 302                   1
..
Wolf                      1
Wyatt Earp                 1
Yankee Zulu                1
Yes, Madam                 1
À la mode                  1
Name: popularity, Length: 1108, dtype: int64

```

```

[36]: # avg runtime
a = df
a = a.groupby('title')['budget'].count()
a

```

```

[36]: title
1-900      1
2 Days in the Valley  1
20,000 Leagues Under the Sea  1
2001: A Space Odyssey  1
301, 302      1
..
Wolf        1
Wyatt Earp  1
Yankee Zulu 1
Yes, Madam  1
À la mode   1
Name: budget, Length: 1108, dtype: int64

```

```

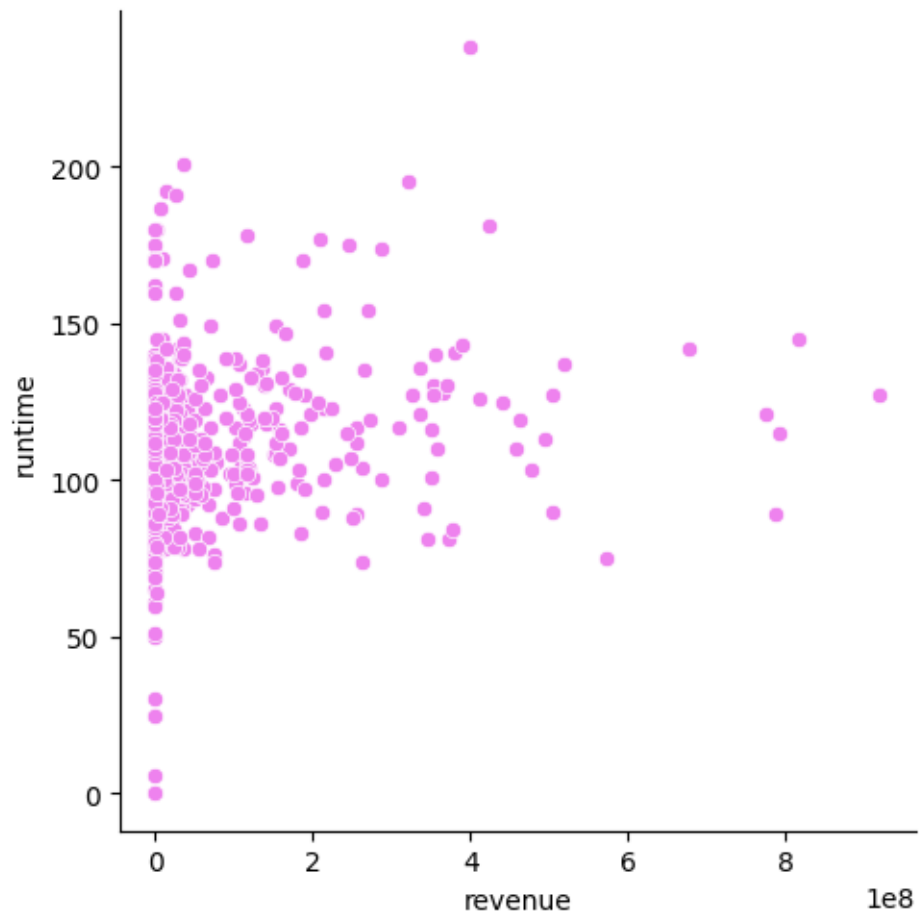
[40]: sns.relplot(data = df , x = 'revenue', y = 'runtime', color = 'violet')

```

```

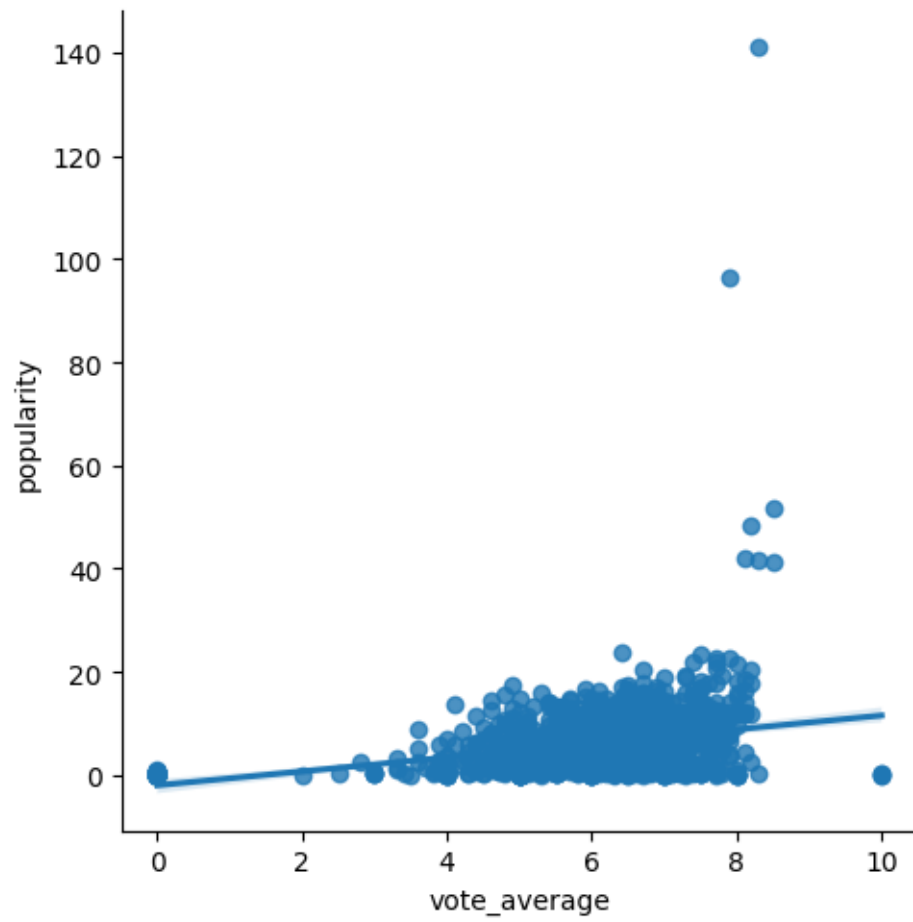
[40]: <seaborn.axisgrid.FacetGrid at 0x7b565e0d6c50>

```



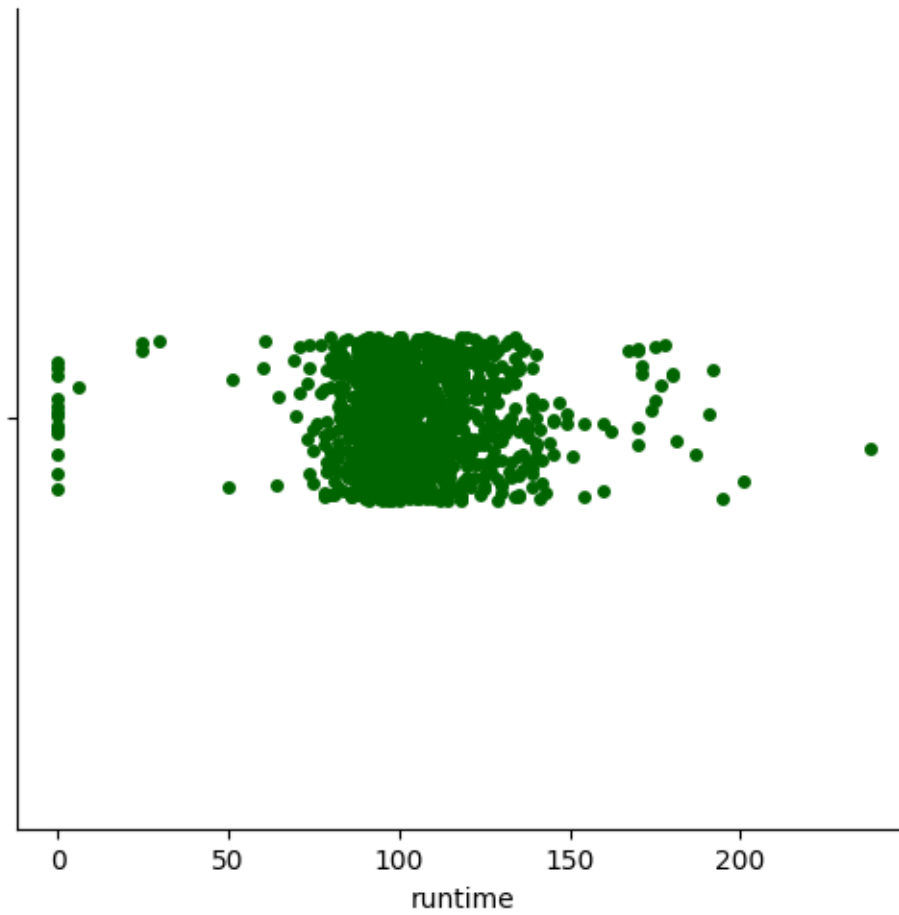
```
[45]: sns.lmplot(data=df, x="vote_average", y="popularity" )
```

```
[45]: <seaborn.axisgrid.FacetGrid at 0x7b565deb41f0>
```



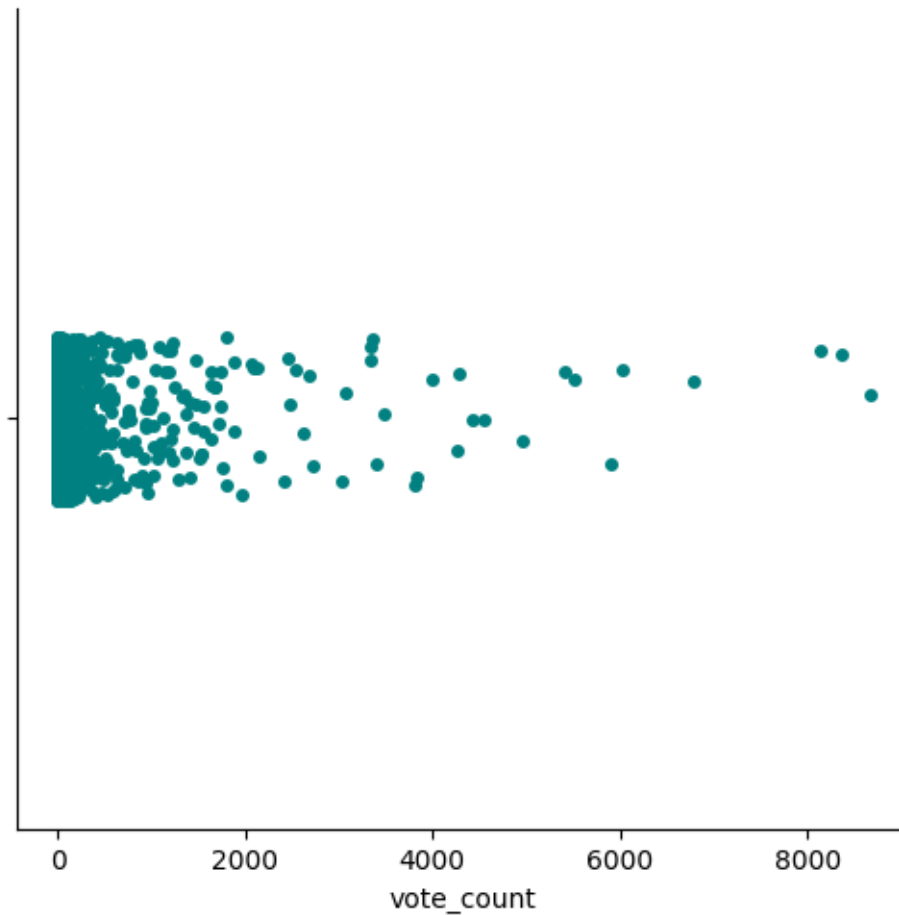
```
[48]: sns.catplot(df , x="runtime" , color='darkgreen')
```

```
[48]: <seaborn.axisgrid.FacetGrid at 0x7b565ded57e0>
```



```
[50]: sns.catplot(df , x="vote_count" , color='teal')
```

```
[50]: <seaborn.axisgrid.FacetGrid at 0x7b565de6bdc0>
```



```
[55]: df['adult'] = df['adult'].replace({'FALSE':1,'RUMORED':0})
df
df['status'] = df['status'].replace({'RELEASED':1})
df
```

```
[55]:
```

	adult	belongs_to_collection	budget	\
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	
1	False		NaN	65000000
2	False	{'id': 119050, 'name': 'Grumpy Old Men Collect...	0	
3	False		NaN	16000000
4	False	{'id': 96871, 'name': 'Father of the Bride Col...	0	
...	
1105	False		NaN	500000
1106	False		NaN	0
1107	False		NaN	0
1108	False		NaN	0
1109	False		NaN	0

```

                                genres \
0      [{'id': 16, 'name': 'Animation'}, {'id': 35, '...
1      [{'id': 12, 'name': 'Adventure'}, {'id': 14, '...
2      [{'id': 10749, 'name': 'Romance'}, {'id': 35, ...
3      [{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam...
4      [{'id': 35, 'name': 'Comedy'}]
...
1105   [{'id': 35, 'name': 'Comedy'}, {'id': 80, 'nam...
1106   [{'id': 99, 'name': 'Documentary'}]
1107   [{'id': 99, 'name': 'Documentary'}]
1108   []
1109   [{'id': 99, 'name': 'Documentary'}]

```

```

                                homepage      id      imdb_id \
0      http://toystory.disney.com/toy-story      862      tt0114709
1      NaN      8844      tt0113497
2      NaN      15602      tt0113228
3      NaN      31357      tt0114885
4      NaN      11862      tt0113041
...
1105   NaN      55049      tt0117968
1106   NaN      124829      tt0116886
1107   NaN      79306      tt0117677
1108   NaN      356054      tt0112759
1109   NaN      10548      tt0118147

```

```

                                original_language      original_title \
0      en      Toy Story
1      en      Jumanji
2      en      Grumpier Old Men
3      en      Waiting to Exhale
4      en      Father of the Bride Part II
...
1105   en      Trois vies et une seule mort
1106   en      The Line King: The Al Hirschfeld Story
1107   de      Snowriders
1108   en      Curtis's Charm
1109   en      When We Were Kings

```

```

                                overview ... release_date \
0      Led by Woody, Andy's toys live happily in his ... ... 10/30/1995
1      When siblings Judy and Peter discover an encha... ... 12/15/1995
2      A family wedding reignites the ancient feud be... ... 12/22/1995
3      Cheated on, mistreated and stepped on, the wom... ... 12/22/1995
4      Just when George Banks has recovered from his ... ... 2/10/1995
...
1105   Raoul Ruiz's enthralling thriller is made up o... ... 5/15/1996

```

```

1106 The childhood, adolescence, and incredible adu... ... 9/27/1996
1107 Spectacular shots of snowcapped peaks and extr... ... 10/17/1996
1108 This film, based on a Jim Carroll short story,... ... 9/18/1995
1109 It's 1974. Muhammad Ali is 32 and thought by m... ... 10/25/1996

    revenue runtime                                spoken_languages \
0      373554033      81.0      [{'iso_639_1': 'en', 'name': 'English'}]
1      262797249     104.0      [{'iso_639_1': 'en', 'name': 'English'}, {'iso...
2           0      101.0      [{'iso_639_1': 'en', 'name': 'English'}]
3      81452156     127.0      [{'iso_639_1': 'en', 'name': 'English'}]
4      76578911     106.0      [{'iso_639_1': 'en', 'name': 'English'}]
...      ...      ...      ...
1105           0     123.0      [{'iso_639_1': 'fr', 'name': 'Français'}]
1106           0      86.0      []
1107     2035470      96.0      [{'iso_639_1': 'en', 'name': 'English'}]
1108           0      74.0      []
1109     2666118      89.0      [{'iso_639_1': 'en', 'name': 'English'}]

    status                                tagline \
0      Released                                NaN
1      Released      Roll the dice and unleash the excitement!
2      Released      Still Yelling. Still Fighting. Still Ready for...
3      Released      Friends are the people who let you be yourself...
4      Released      Just When His World Is Back To Normal... He's ...
...      ...      ...
1105      Released                                NaN
1106      Released                                NaN
1107      Released      Man Wasn't Meant To Fly... Until Now!
1108      Released                                NaN
1109      Released      The untold story of the Rumble in the Jungle.

                                title  video  vote_average  vote_count
0                                Toy Story  False           7.7         5415
1                                Jumanji  False           6.9         2413
2                                Grumpier Old Men  False           6.5           92
3                                Waiting to Exhale  False           6.1           34
4                                Father of the Bride Part II  False           5.7          173
...      ...      ...      ...
1105      Three Lives and Only One Death  False           8.0           5
1106      The Line King: The Al Hirschfeld Story  False           0.0           0
1107                                Snowriders  False           0.0           0
1108                                Curtis's Charm  False           0.0           0
1109                                When We Were Kings  False           7.3          54

```

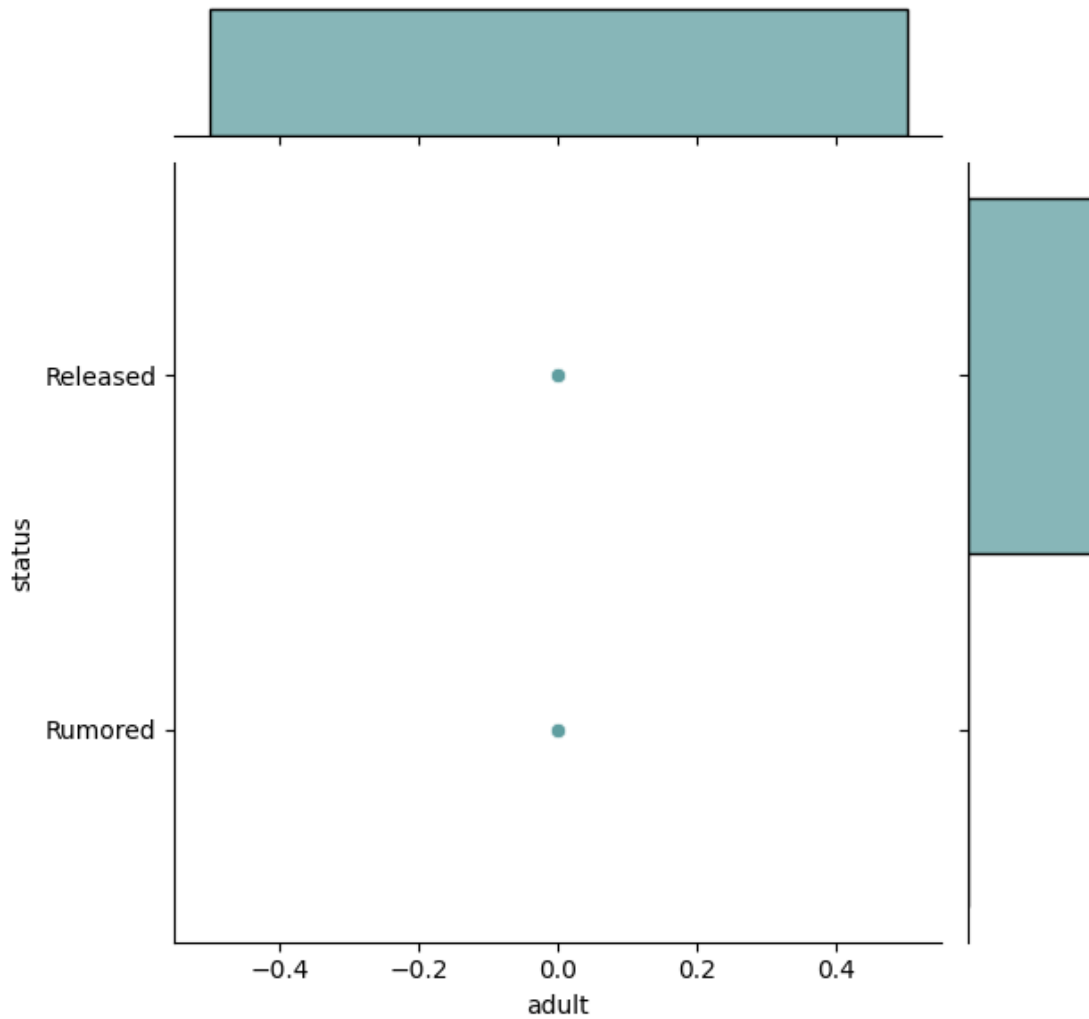
[1110 rows x 24 columns]

```
[56]: sns.jointplot(data = df , x = 'adult' , y='status', color = 'cadetblue')
```



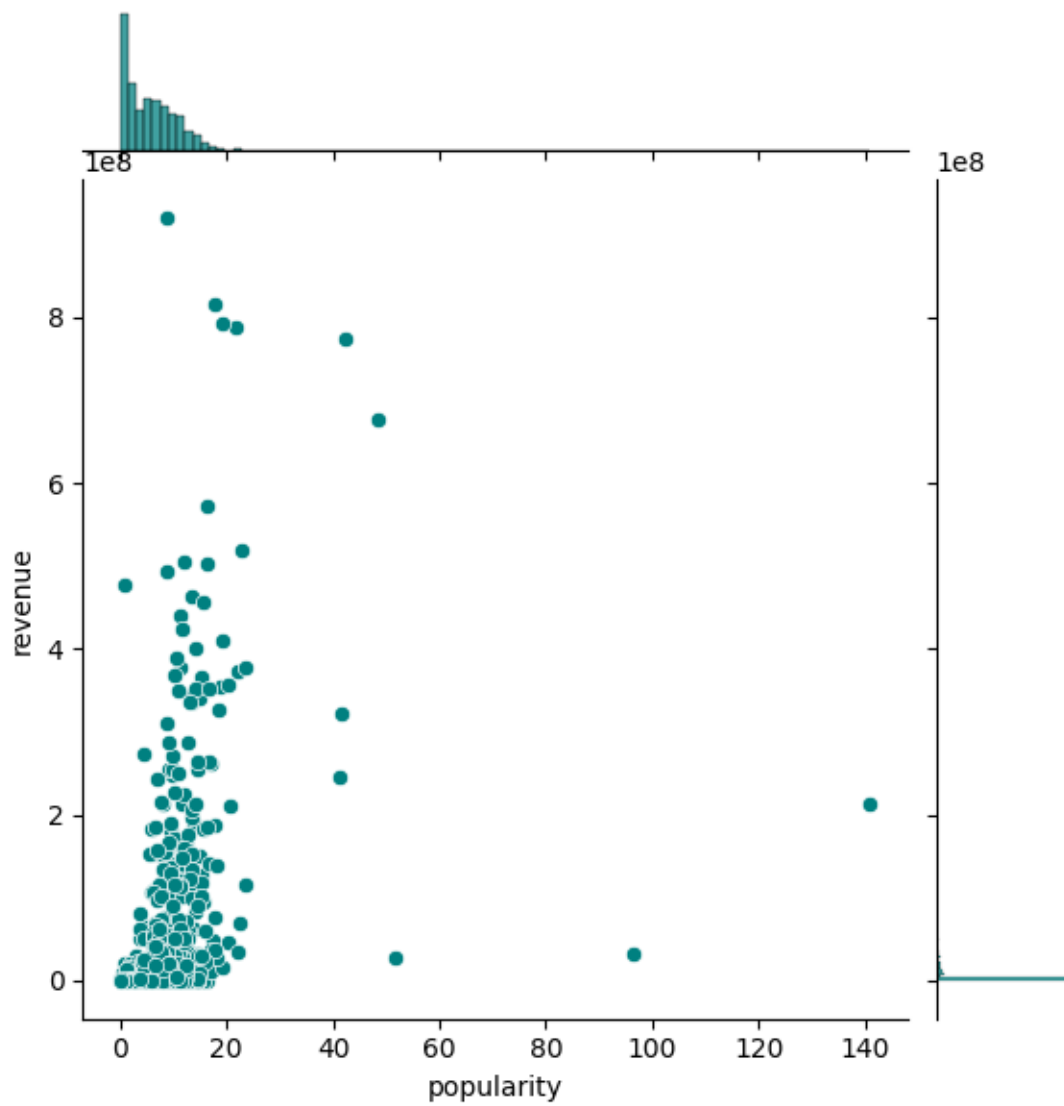
```
<__array_function__ internals>:180: RuntimeWarning: Converting input from bool  
to <class 'numpy.uint8'> for compatibility.
```

```
[56]: <seaborn.axisgrid.JointGrid at 0x7b565d2715d0>
```



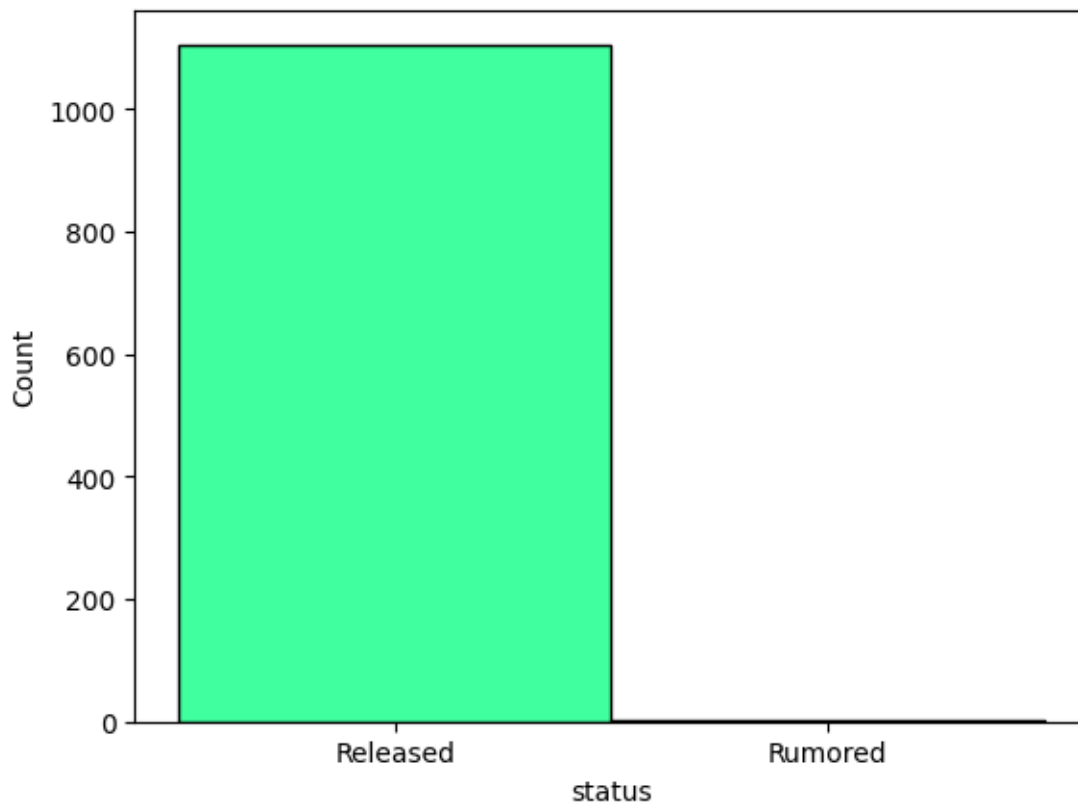
```
[53]: sns.jointplot(data = df , x = 'popularity' , y='revenue', color = 'teal')
```

```
[53]: <seaborn.axisgrid.JointGrid at 0x7b565d637a00>
```



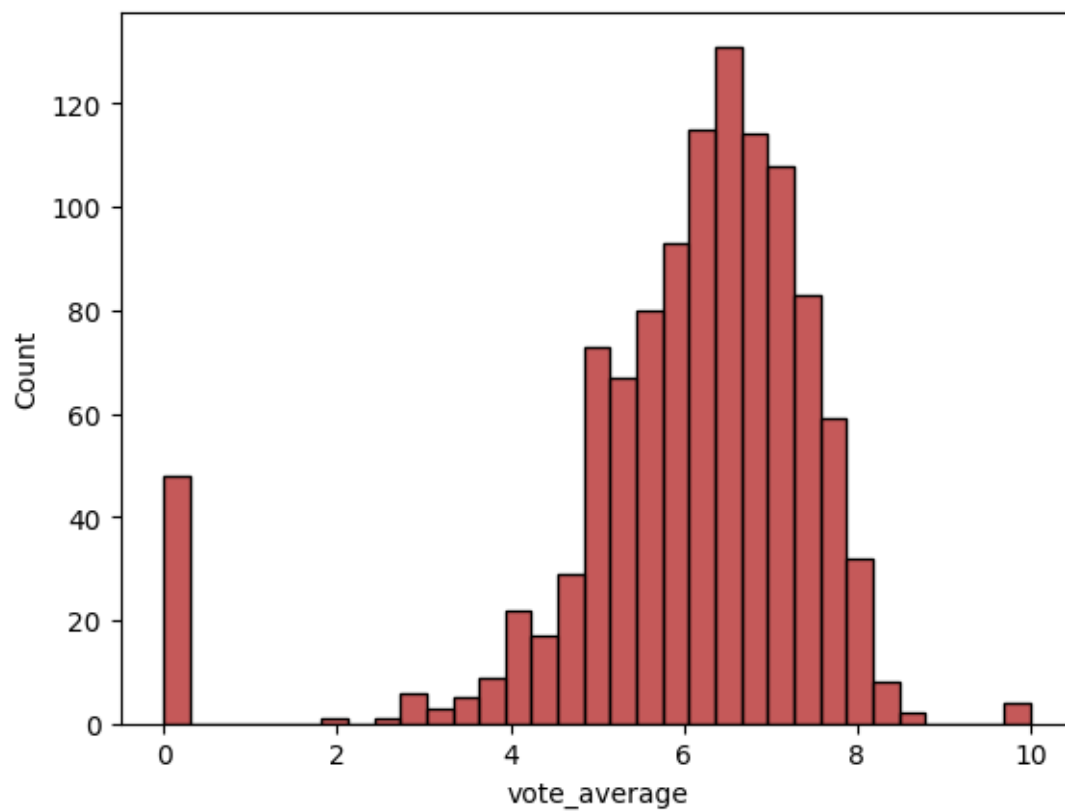
```
[57]: sns.histplot(df , x="status" , color='springgreen')
```

```
[57]: <Axes: xlabel='status', ylabel='Count'>
```



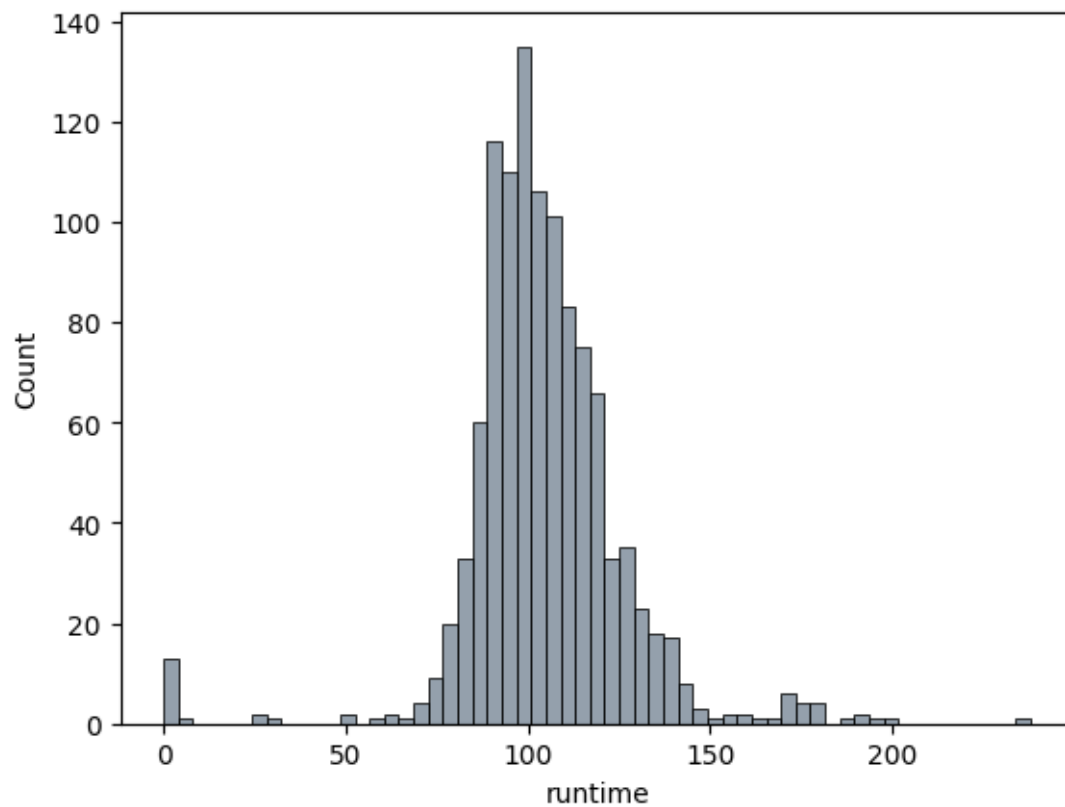
```
[58]: sns.histplot(df , x="vote_average" , color='firebrick')
```

```
[58]: <Axes: xlabel='vote_average', ylabel='Count'>
```



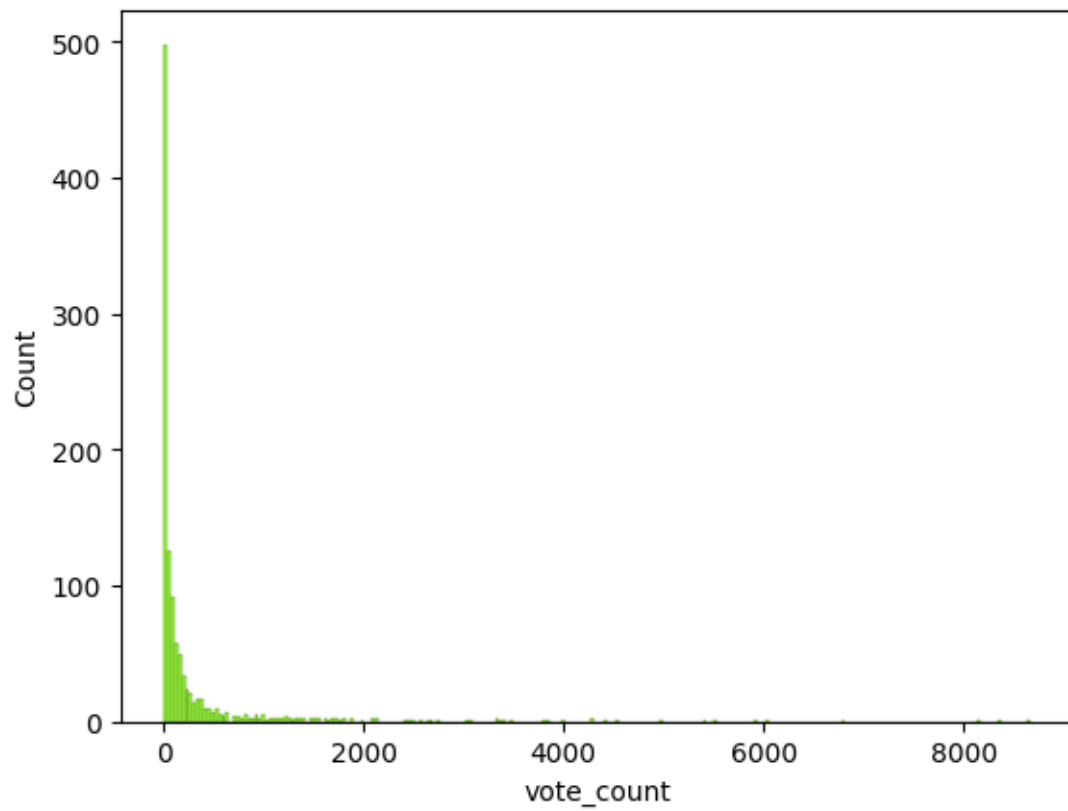
```
[59]: sns.histplot(df , x="runtime" , color='slategray')
```

```
[59]: <Axes: xlabel='runtime', ylabel='Count'>
```



```
[60]: sns.histplot(df , x="vote_count" , color='lawngreen')
```

```
[60]: <Axes: xlabel='vote_count', ylabel='Count'>
```



```
[61]: sns.boxenplot(df , x="budget" , y='revenue', color = 'deeppink')
```

```
[61]: <Axes: xlabel='budget', ylabel='revenue'>
```

