FINAL PROJECT:
Top 10000 Popular Movies Dataset

https://www.kaggle.com/datasets/omkarborikar/top-10000-popular-movies

Clean the Dataset :

	Unnamed: 0	id	original_language	original_title	popularity	release_date	vote_average	vote_count	genre	overview	revenue	runtime	tagline
0	0	580489.0	en	Venom: Let There Be Carnage	5401.308	2021-09-30	6.8	1736.0	['Science Fiction', 'Action', 'Adventure']	After finding a host body in investigative rep...	424000000.0	97.0	NaN
1	1	524434.0	en	Eternals	3365.535	2021-11-03	7.1	622.0	['Action', 'Adventure', 'Science Fiction', 'Fa...	The Eternals are a team of ancient aliens who ...	165000000.0	157.0	In the beginning...
2	2	438631.0	en	Dune	2911.423	2021-09-15	8.0	3632.0	['Action', 'Adventure', 'Science Fiction']	Paul Atreides, a brilliant and gifted young ma...	331116356.0	155.0	Beyond fear, destiny awaits.
3	3	796499.0	en	Army of Thieves	2552.437	2021-10-27	6.9	555.0	['Action', 'Crime', 'Thriller']	A mysterious woman recruits bank teller Ludwig...	0.0	127.0	Before Vegas, one locksmith became a legend.
4	4	550988.0	en	Free Guy	1850.470	2021-08-11	7.8	3493.0	['Comedy', 'Action', 'Adventure', 'Science Fic...	A bank teller called Guy realizes he is a back...	331096766.0	115.0	Life's too short to be a background character.
...	...	...	...	...	...	...	...	...	...	...	...	...	...
10009	9995	530.0	en	A Grand Day Out	9.266	1990-05-18	7.5	594.0	['Adventure', 'Animation', 'Comedy', 'Science ...	Wallace and Gromit have run out of cheese and ...	0.0	23.0	NaN
10010	9996	15934.0	en	El cantante	10.417	2006-09-12	7.0	80.0	['History', 'Drama', 'Music']	The rise and fall of salsa singer, Héctor Lavo...	0.0	116.0	Based on the true story of the King of Salsa, ...
10011	9997	162215.0	en	How I Live Now	9.520	2013-09-10	6.6	705.0	['Drama', 'Action', 'Thriller', 'War']	An American girl on holiday in the English cou...	0.0	101.0	Love will lead you home
10012	9998	5723.0	en	Once	9.267	2007-03-23	7.4	990.0	['Drama', 'Music', 'Romance']	A vacuum repairman moonlights as a street musi...	20710513.0	85.0	How often do you find the right person?
10013	9999	311667.0	en	Manhattan Night	9.273	2016-05-20	6.0	304.0	['Drama', 'Mystery', 'Thriller']	Porter Wren is a Manhattan tabloid writer with...	0.0	113.0	No reporter can resist a siren.

- ~~id - Every movie has its unique ID.~~
- **original_language** - There are total 44 languages present in this column. Total 7771 movies with 'English' as original language. Values in this column are ISO 639-1 codes of languages. I.e 'en' for 'English' , 'hi' for 'Hindi' etc.
- **original_title** - Title of the movie.
- **popularity** - Popularity of movie. Bigger the number , higher the popularity.
- **release_date** - Release date of the movie. If release date is not present for any movie , then that movie is not released yet.
- **vote_average** - Average of rating/vote for the movie.
- **vote_count** - Number of ratings/vote recorded for the movie.
- **genre** - Genre of the movie.
- ~~overview - Brief description of movie in string format.~~
- **revenue** - Revenue of Movie
- **runtime** - Runtime of movie in minutes.
- ~~tagline - Tagline of the movie~~
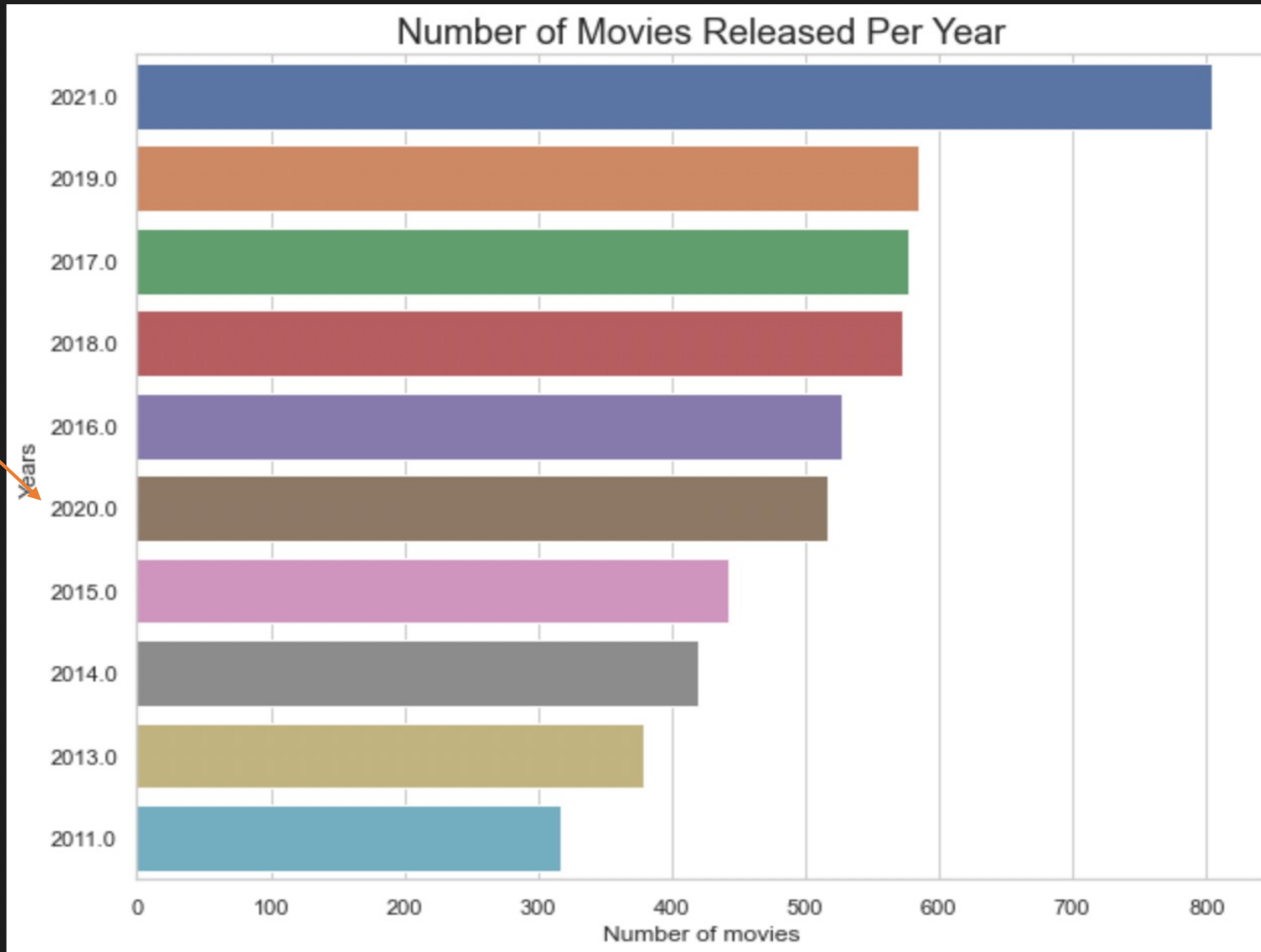
Objectives:
- Best Movie to Produce

```python
1  df
```

[67] ✓ 0.2s   Python

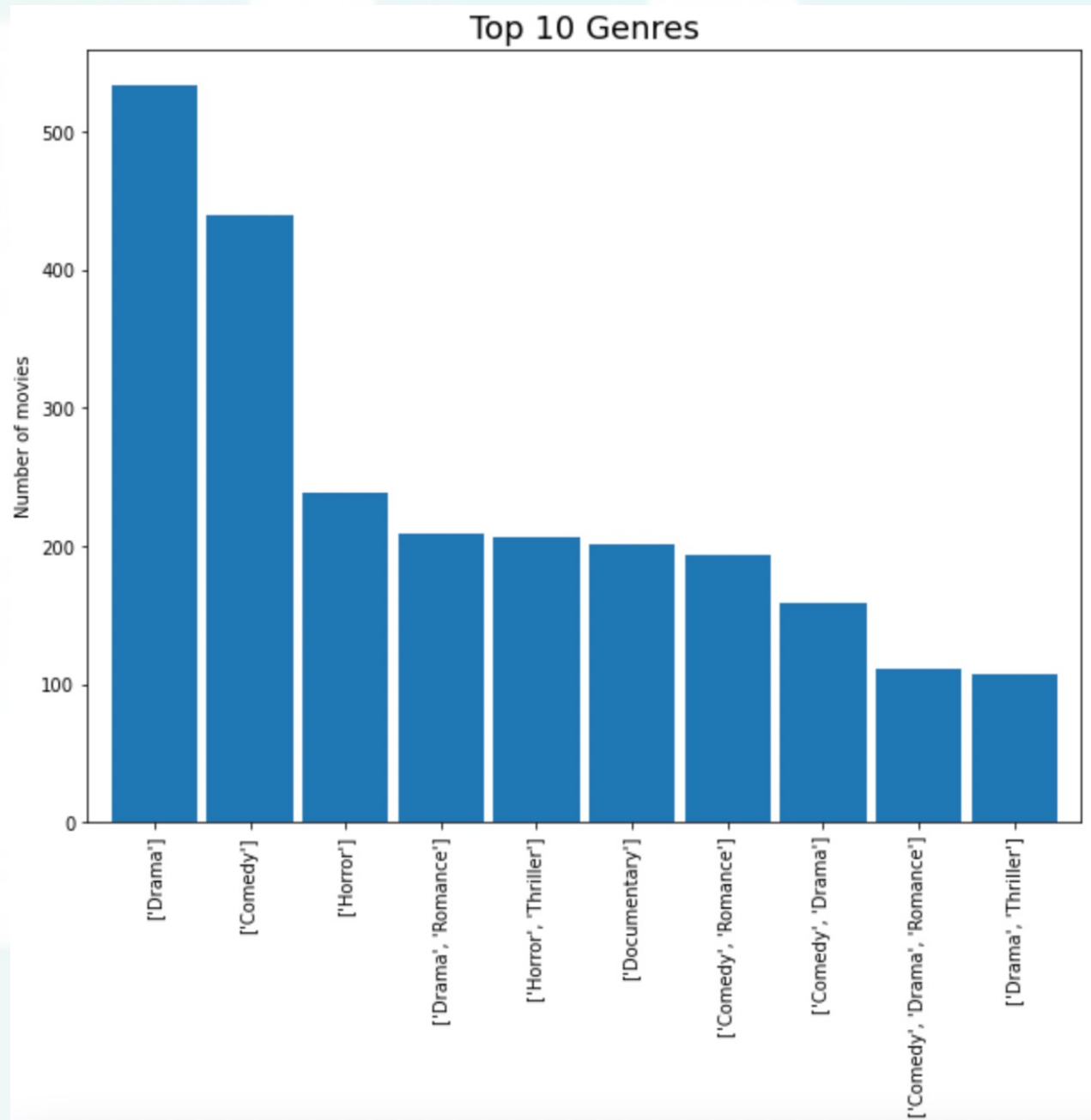|	original_language	original_title	popularity	release_date	vote_average	vote_count	genre	revenue	runtime
0	en	Venom: Let There Be Carnage	5401.308	2021.0	6.8	1736.0	['Science Fiction', 'Action', 'Adventure']	424000000.0	97.0
1	en	Eternals	3365.535	2021.0	7.1	622.0	['Action', 'Adventure', 'Science Fiction', 'Fa...	165000000.0	157.0
2	en	Dune	2911.423	2021.0	8.0	3632.0	['Action', 'Adventure', 'Science Fiction']	331116356.0	155.0
3	en	Army of Thieves	2552.437	2021.0	6.9	555.0	['Action', 'Crime', 'Thriller']	0.0	127.0
4	en	Free Guy	1850.470	2021.0	7.8	3493.0	['Comedy', 'Action', 'Adventure', 'Science Fic...	331096766.0	115.0
...	...	...	...	...	...	...	...	...	...
10009	en	A Grand Day Out	9.266	1990.0	7.5	594.0	['Adventure', 'Animation', 'Comedy', 'Science ...	0.0	23.0
10010	en	El cantante	10.417	2006.0	7.0	80.0	['History', 'Drama', 'Music']	0.0	116.0
10011	en	How I Live Now	9.520	2013.0	6.6	705.0	['Drama', 'Action', 'Thriller', 'War']	0.0	101.0
10012	en	Once	9.267	2007.0	7.4	990.0	['Drama', 'Music', 'Romance']	20710513.0	85.0
10013	en	Manhattan Night	9.273	2016.0	6.0	304.0	['Drama', 'Mystery', 'Thriller']	0.0	113.0

9867 rows × 9 columns

a) Number of Movies Released per Year
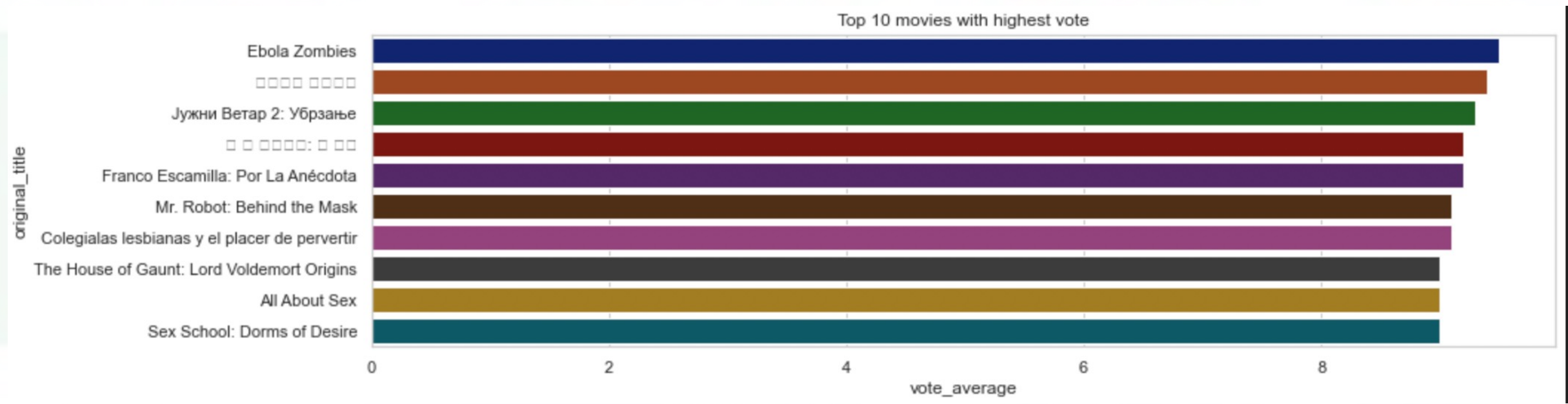b) Does Covid affect the number of Movies released?
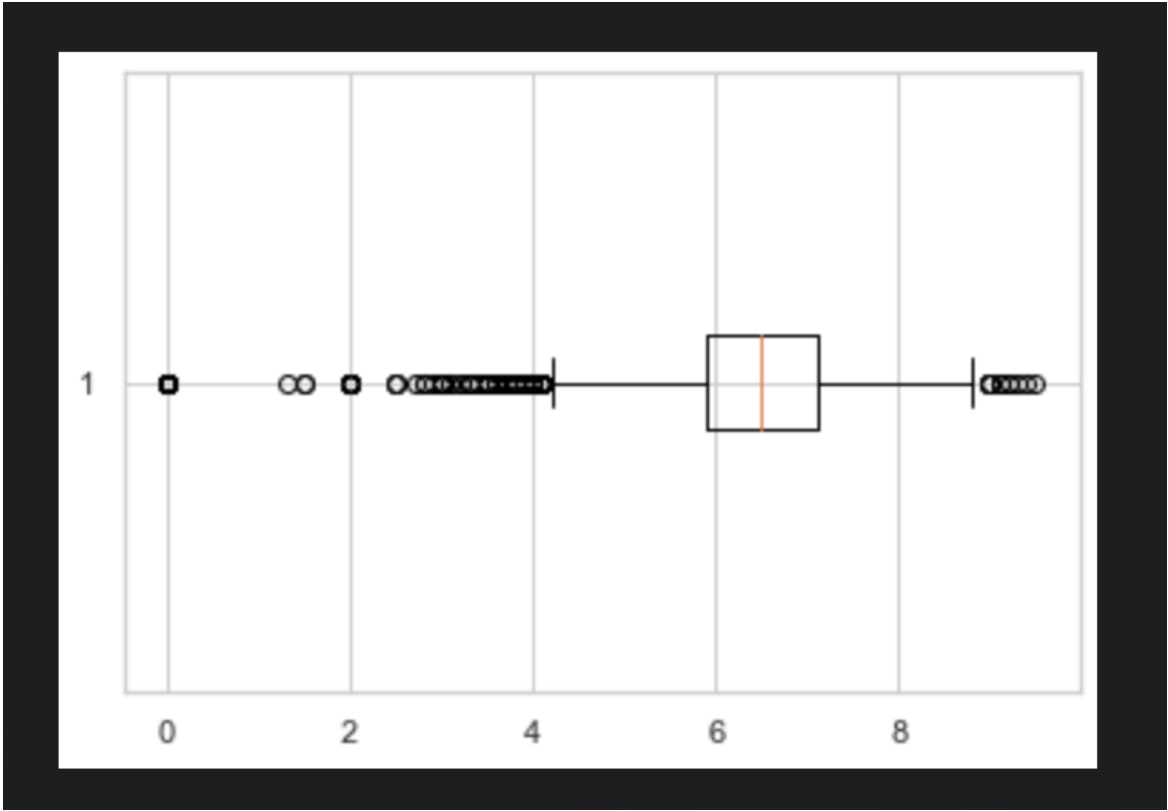
Finding the top 10 Genre:

Top 10 Genres
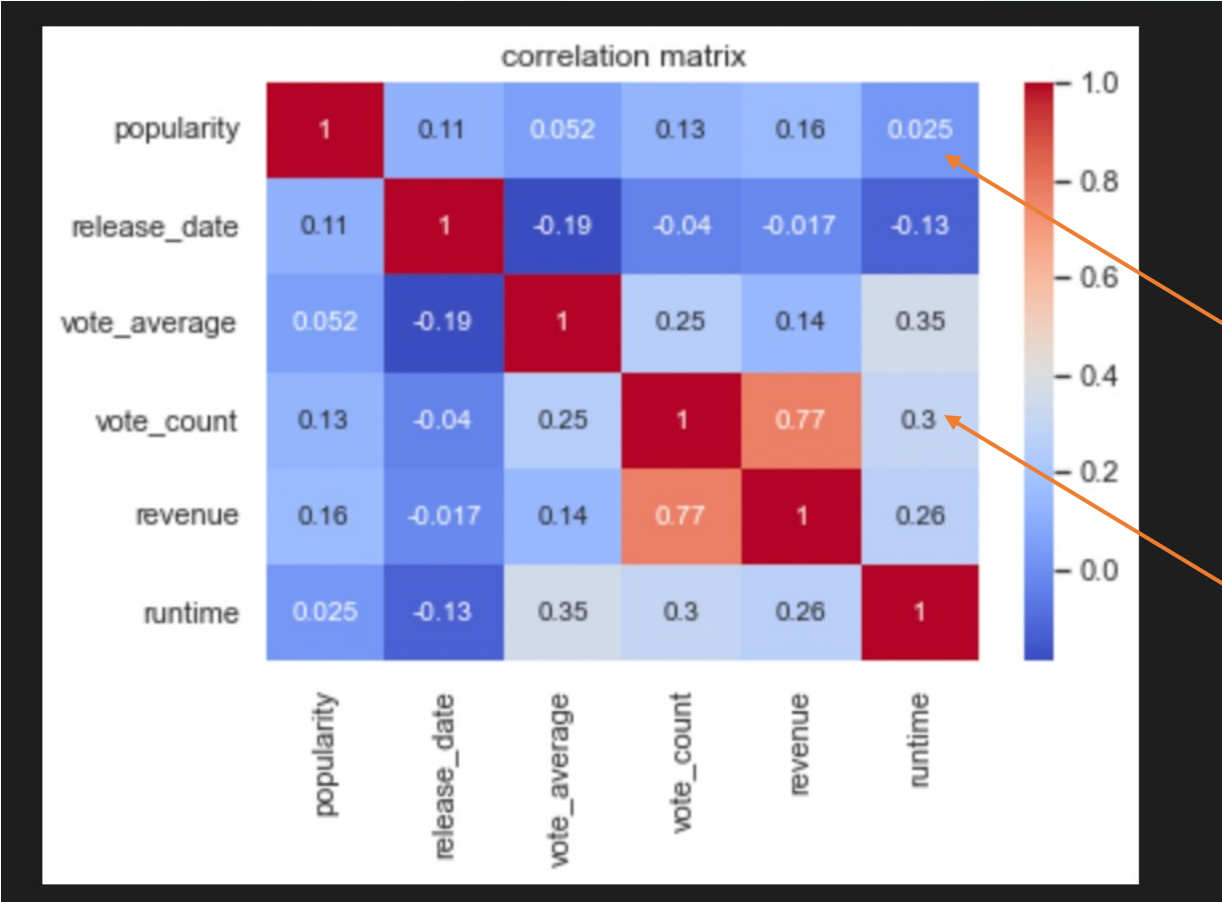
Finding the top 10 movies with the highest vote:
Is it Valid?

Top 10 movies with highest vote

	original_language	original_title	popularity	release_date	vote_average	vote_count	genre
5024	zh	Ebola Zombies	14.316	2015.0	9.5	2.0	['Horror', 'Science Fiction']

Outliers for Rating:

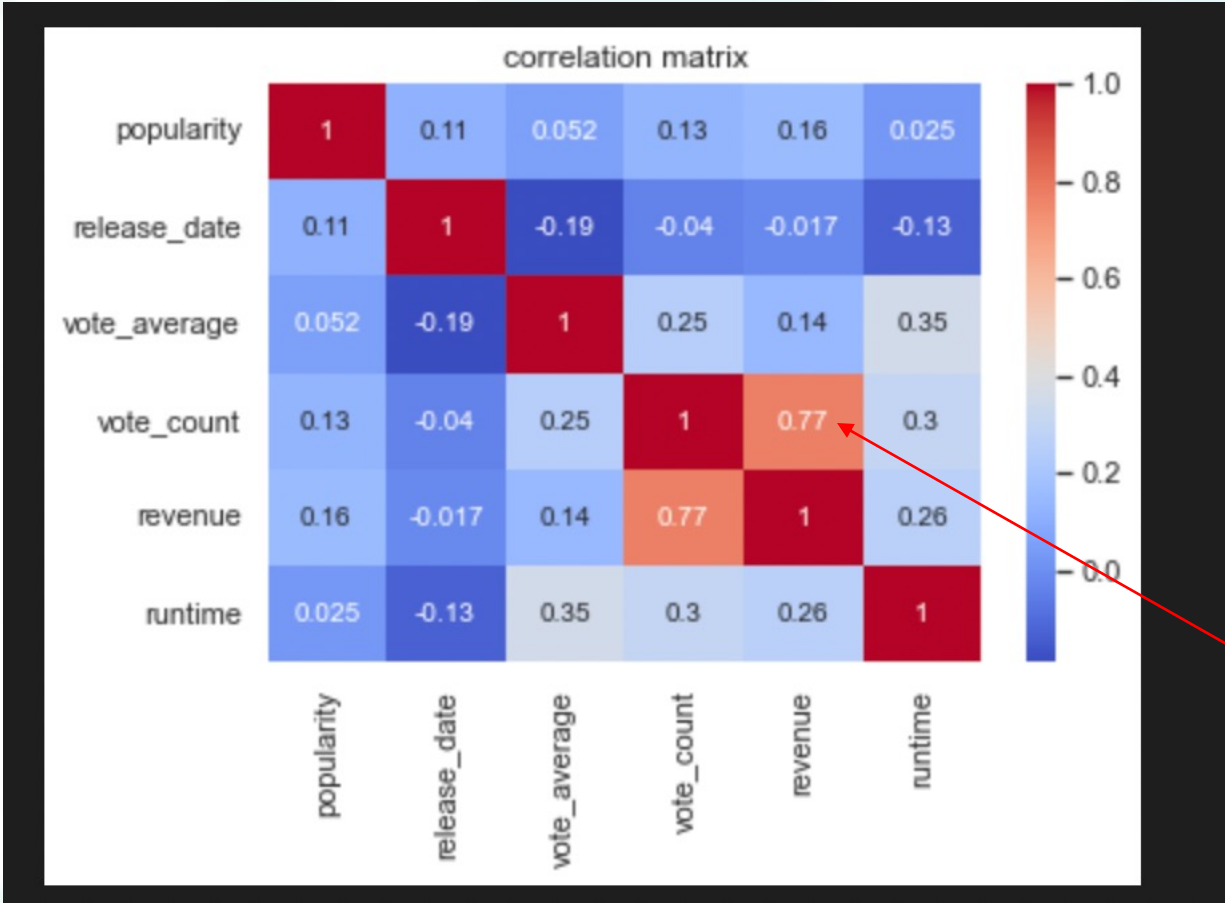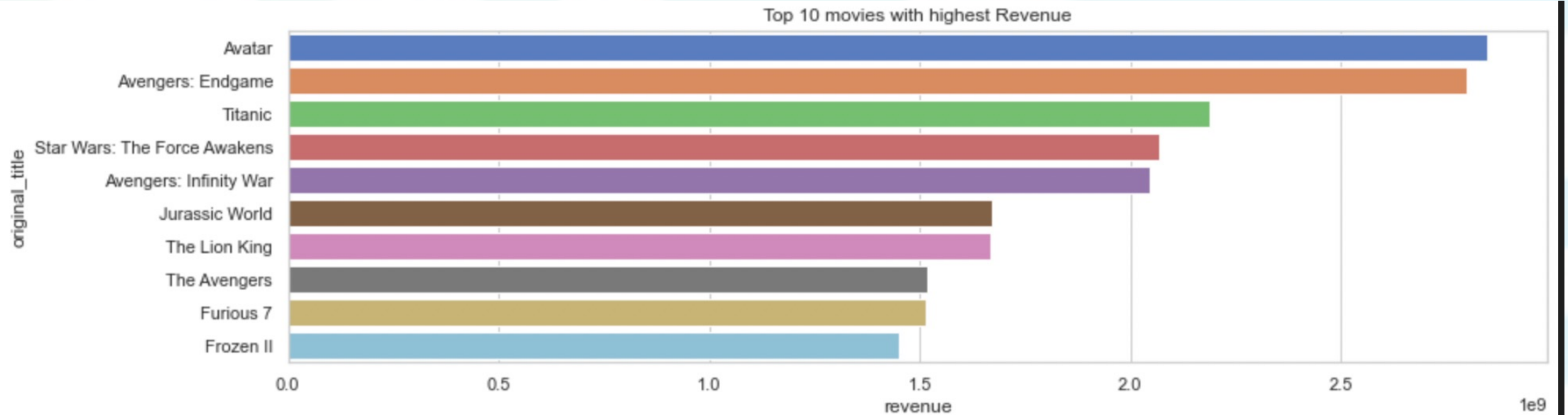Is there any correlation between the runtime and vote count/popularity:

correlation matrix

	popularity	release_date	vote_average	vote_count	revenue	runtime
popularity	1	0.11	0.052	0.13	0.16	0.025
release_date	0.11	1	-0.19	-0.04	-0.017	-0.13
vote_average	0.052	-0.19	1	0.25	0.14	0.35
vote_count	0.13	-0.04	0.25	1	0.77	0.3
revenue	0.16	-0.017	0.14	0.77	1	0.26
runtime	0.025	-0.13	0.35	0.3	0.26	1

```
1  df.describe()
✓ 0.3s
```

	popularity	release_date	vote_average	vote_count	revenue	runtime
count	9818.000000	9818.000000	9818.000000	9818.000000	9.816000e+03	9813.000000
mean	34.758134	2007.475963	6.364575	1336.219291	5.827067e+07	99.869968
std	101.577757	14.657145	1.295426	2517.924189	1.492031e+08	27.096489

Is there any correlation between there revenue and vote count:

Finding the top 10 movies with the highest revenue:
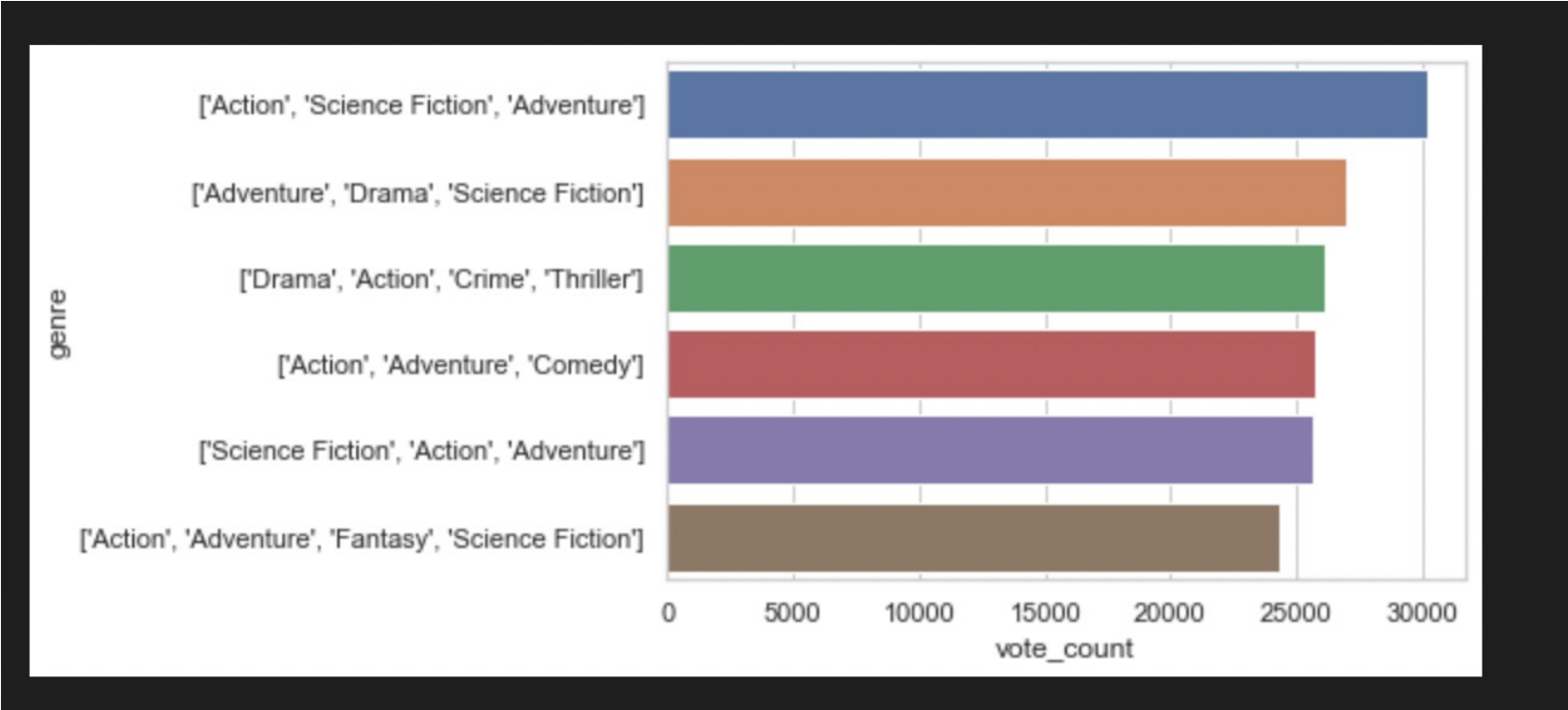[Action, Adventure, Science Fiction]

Top 10 movies with highest Revenue

```
1
2  df.loc[df["original_title"] =="Avatar"]
3
4
```
✓ 0.5s

	original_language	original_title	popularity	release_date	vote_average	vote_count	genre	revenue	runtime
66	en	Avatar	374.059	2009.0	7.5	24280.0	['Action', 'Adventure', 'Fantasy', 'Science Fi...	2.847246e+09	162.0

[+ Code] [+ Markdown]

```
1  df.loc[df["original_title"] =="Avengers: Endgame"]
```
✓ 0.3s

	original_language	original_title	popularity	release_date	vote_average	vote_count	genre	revenue	runtime
167	en	Avengers: Endgame	193.86	2019.0	8.3	19434.0	['Adventure', 'Science Fiction', 'Action']	2.797801e+09	181.0

Finding the top 6 Genre with highest vote count:

So is it true that Movies with [Action/ Adventure/ Science] Fiction have high votes?