

Microsoft Movie studio project

Analysis by

Alireza Karimi

September 11th, 2020



Overview:

- Start to produce movies

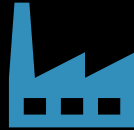
Top 20 Genres

	Genres	Number of Rating
0	Drama	10189
1	Documentary	9149
2	Comedy	4604
3	Comedy,Drama	2451
4	Horror	2232
5	Drama,Romance	1392
6	Thriller	1288
7	Comedy,Drama,Romance	1159
8	Comedy,Romance	1128
9	Horror,Thriller	943
10	Drama,Thriller	938
11	Action	721
12	Biography,Documentary	675
13	Documentary,Drama	572
14	Documentary,Music	544
15	Action,Crime,Drama	530
16	Romance	529
17	Comedy,Horror	515
18	Crime,Drama,Thriller	493
19	Biography,Documentary,History	476

Business problem:



Hard to be part of film
industry



New to the industry



Need to know what to
make

Input Data

- 2 data tables from IMDB
- 1 data table from Box office mojo

Raw tables from
our resources



IMDB primaryTable

	tconst	primary_title	original_title	start_year	runtime_minutes	genres
0	tt0063540	Sunghursh	Sunghursh	2013	175.0	Action,Crime,Drama
1	tt0066787	One Day Before the Rainy Season	Ashad Ka Ek Din	2019	114.0	Biography,Drama
2	tt0069049	The Other Side of the Wind	The Other Side of the Wind	2018	122.0	Drama
3	tt0069204	Sabse Bada Sukh	Sabse Bada Sukh	2018	NaN	Comedy,Drama
4	tt0100275	The Wandering Soap Opera	La Telenovela Errante	2017	80.0	Comedy,Drama,Fantasy
...
146139	tt9916538	Kuambil Lagi Hatiku	Kuambil Lagi Hatiku	2019	123.0	Drama
146140	tt9916622	Rodolpho Teóphilo - O Legado de um Pioneiro	Rodolpho Teóphilo - O Legado de um Pioneiro	2015	NaN	Documentary
146141	tt9916706	Dankyavar Danka	Dankyavar Danka	2013	NaN	Comedy
146142	tt9916730	6 Gunn	6 Gunn	2017	116.0	NaN
146143	tt9916754	Chico Albuquerque - Revelações	Chico Albuquerque - Revelações	2013	NaN	Documentary

IMDB Average Rating Table

	tconst	averagerating	numvotes
0	tt10356526	8.3	31
1	tt10384606	8.9	559
2	tt1042974	6.4	20
3	tt1043726	4.2	50352
4	tt1060240	6.5	21
...
73851	tt9805820	8.1	25
73852	tt9844256	7.5	24
73853	tt9851050	4.7	14
73854	tt9886934	7.0	5
73855	tt9894098	6.3	128

Data preparation

With Python
and some
libraries:

- Pandas
- Matplotlib
- Numpy
- Seaborn

Merging Datasets

Merging tables to get data to make
analysis



	tconst	averagerating	numvotes	primary_title	original_title	start_year	runtime_minutes	genres
0	tt10356526	8.3	31	Laiye Je Yaarian	Laiye Je Yaarian	2019	117.0	Romance
1	tt10384606	8.9	559	Borderless	Borderless	2019	87.0	Documentary
2	tt1042974	6.4	20	Just Inès	Just Inès	2010	90.0	Drama
3	tt1043726	4.2	50352	The Legend of Hercules	The Legend of Hercules	2014	99.0	Action,Adventure,Fantasy
4	tt1060240	6.5	21	Até Onde?	Até Onde?	2011	73.0	Mystery,Thriller
...
65715	tt9768966	8.6	27	Plugged in	Plugged in	2019	53.0	Documentary
65716	tt9805820	8.1	25	Caisa	Caisa	2018	84.0	Documentary
65717	tt9844256	7.5	24	Code Geass: Lelouch of the Rebellion - Glorifi...	Code Geass: Lelouch of the Rebellion Episode III	2018	120.0	Action,Animation,Sci-Fi
65718	tt9886934	7.0	5	The Projectionist	The Projectionist	2019	81.0	Documentary
65719	tt9894098	6.3	128	Sathru	Sathru	2019	129.0	Thriller

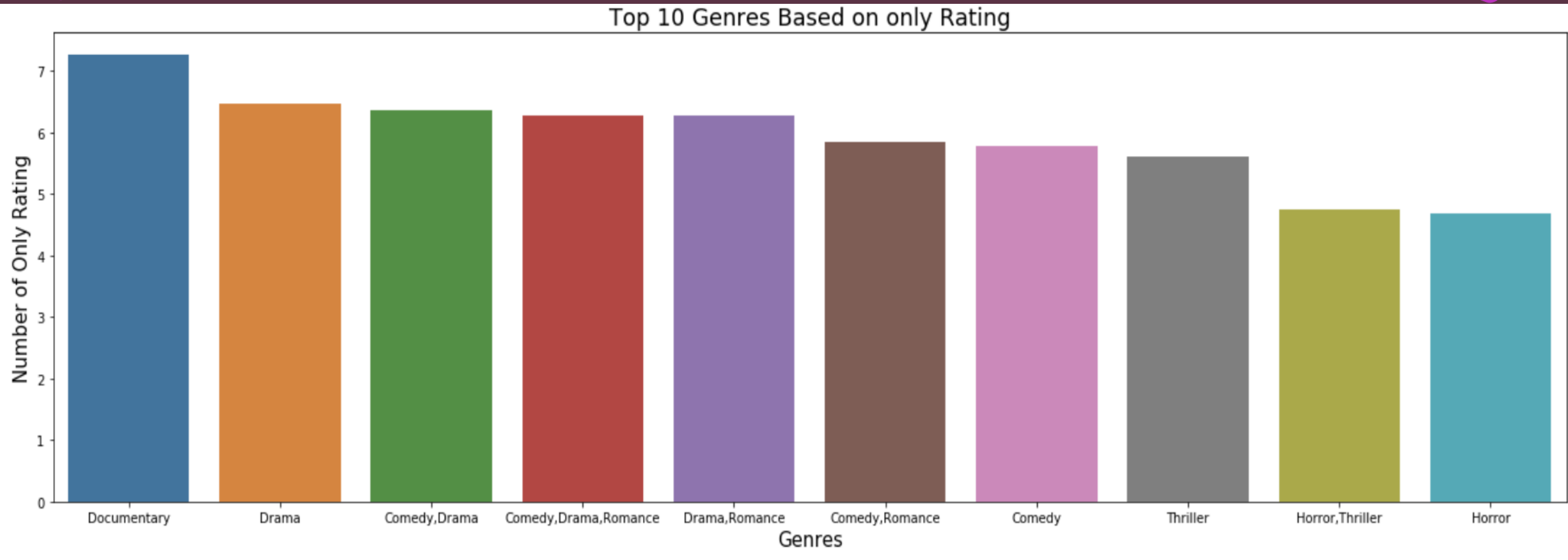
	tconst	averagerating	numvotes	primary_title	original_title	runtime_minutes	genres	rating*numvotes
0	tt10356526	8.3	31	Laiye Je Yaarian	Laiye Je Yaarian	117.0	Romance	257.3
1	tt10384606	8.9	559	Borderless	Borderless	87.0	Documentary	4975.1
2	tt1042974	6.4	20	Just Inès	Just Inès	90.0	Drama	128.0
3	tt1043726	4.2	50352	The Legend of Hercules	The Legend of Hercules	99.0	Action,Adventure,Fantasy	211478.4
4	tt1060240	6.5	21	Até Onde?	Até Onde?	73.0	Mystery,Thriller	136.5
...
65715	tt9768966	8.6	27	Plugged in	Plugged in	53.0	Documentary	232.2
65716	tt9805820	8.1	25	Caisa	Caisa	84.0	Documentary	202.5
65717	tt9844256	7.5	24	Code Geass: Lelouch of the Rebellion - Glorifi...	Code Geass: Lelouch of the Rebellion Episode III	120.0	Action,Animation,Sci-Fi	180.0
65718	tt9886934	7.0	5	The Projectionist	The Projectionist	81.0	Documentary	35.0
65719	tt9894098	6.3	128	Sathru	Sathru	129.0	Thriller	806.4

65720 rows × 8 columns

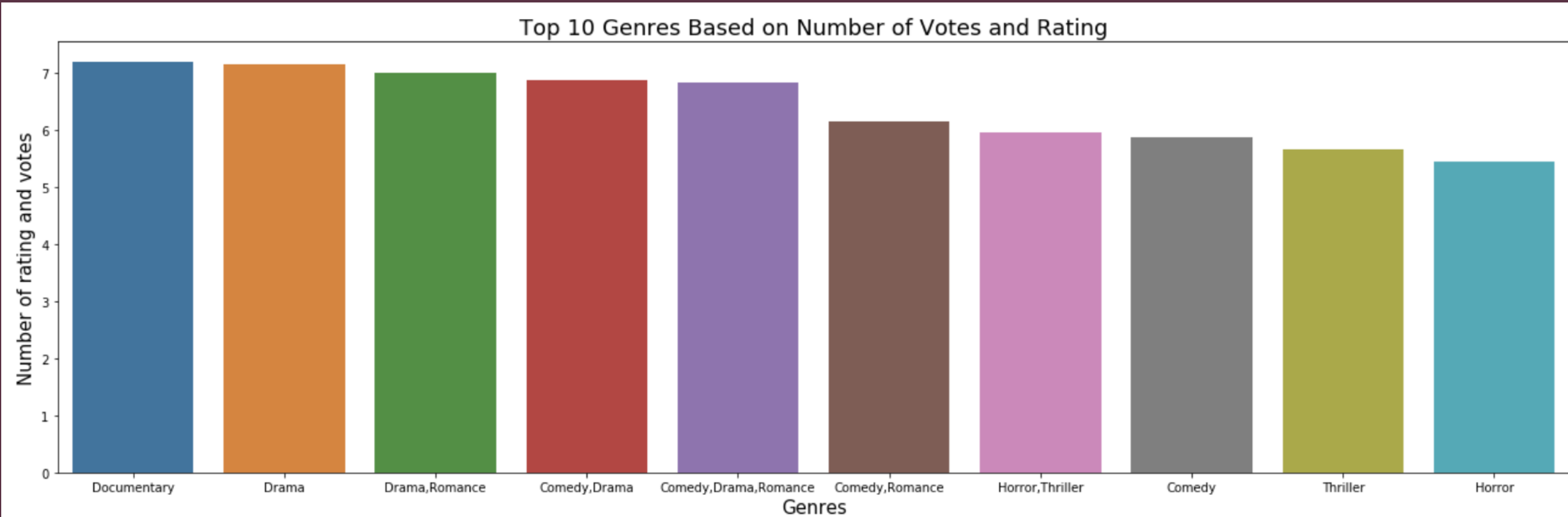


Analysis

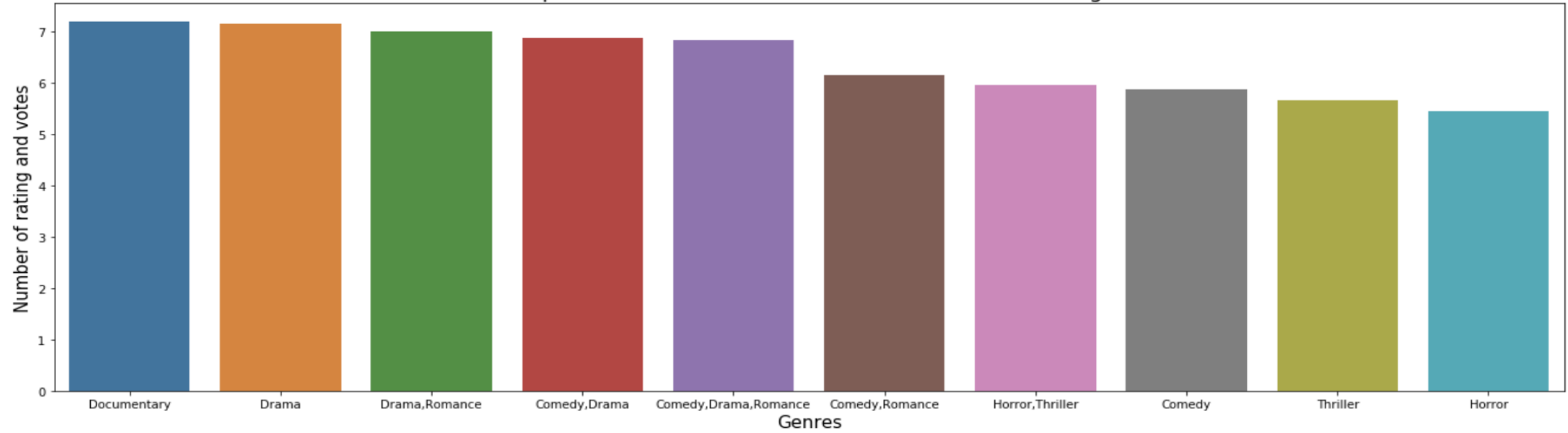
Top 10 Genres Based on Only Rating



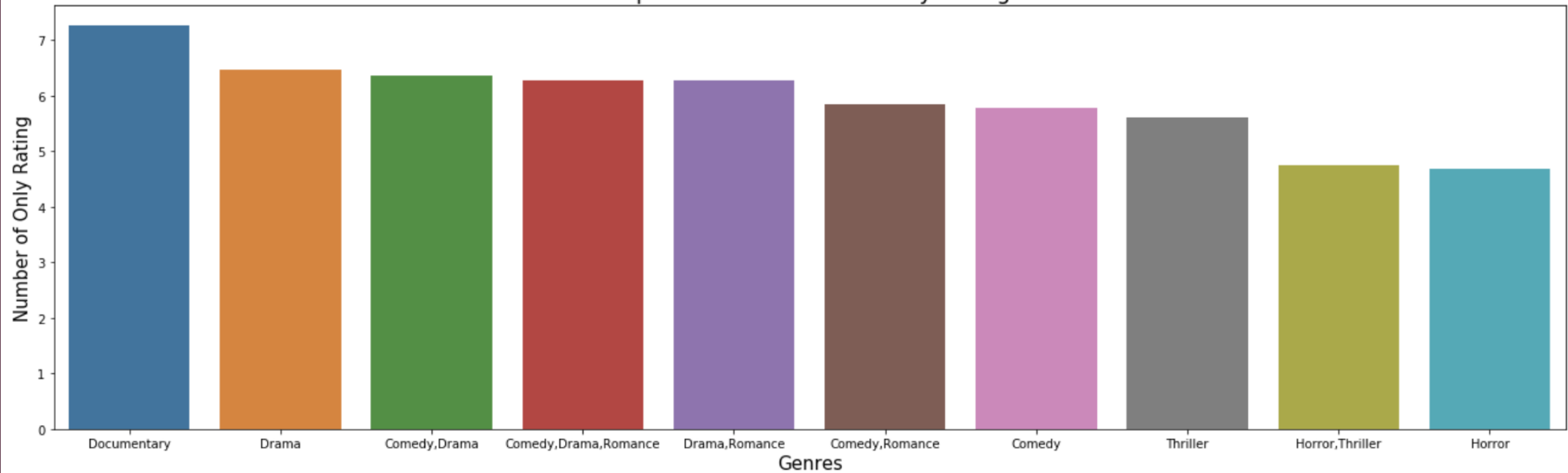
Top 10 Genres Based on Number of Votes and Rating



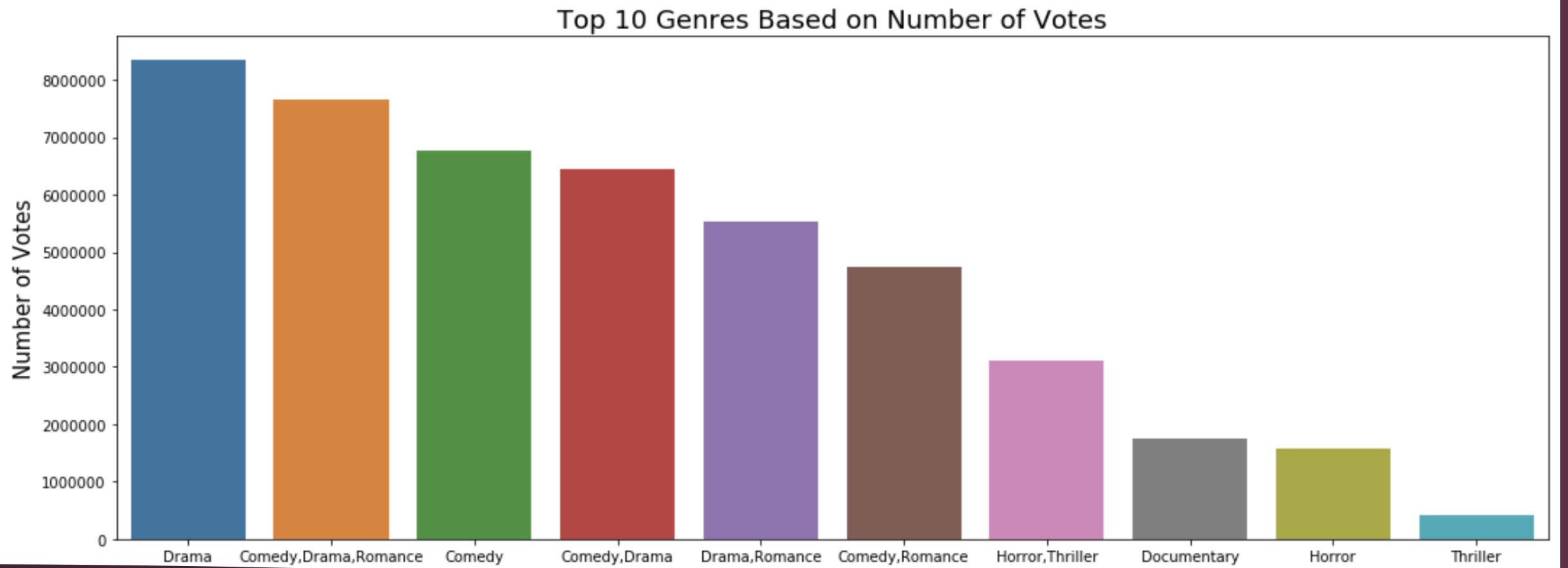
Top 10 Genres Based on Number of Votes and Rating



Top 10 Genres Based on only Rating



Top 10 Genres Based on Number of Votes





Conclusions

If they really want to make movie, they'd
better to make ...

Next step

They need to negotiate with some good movie makers to get the job done..,

But



An Extra Point:

Top 10 Runing Time in Minute

