

# IMDb Movie Analysis

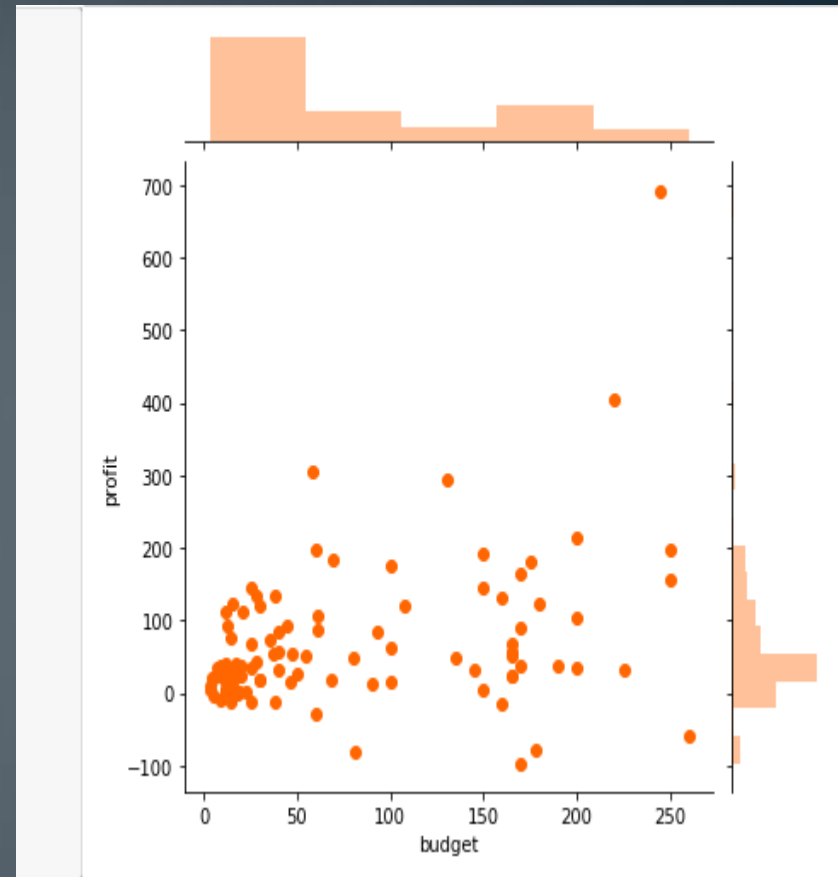


# Data Exploration

- Data set contains 100 rows and 62 columns.
- **Problem Statement:** Finding out insights about the movies and their voters and suggest way to improve.
- Based on the problem statement, we should focus on the voters and the genres.

# Let's Talk Profit!

- Created a column called “Profit” which contains the difference of gross and budget.
- Plotted a scatter plot between the columns budget and profit.
- Individually inspecting budget and profit, both are slightly skewed to right, which means most of the movies have positive profit and also budgets are high.



# The General Audience and the Critics

- Converted the Metacritic column to a scale of 10 and then created a new column called as “Avg\_rating” which will have the average of the MetaCritic and IMDbrating.
- Retained only the movies with a rating equal to or higher than 8 and stored these movies in a new dataframe “UniversalAcclaim”.

```
In [18]: UniversalAcclaim
```

```
Out[18]:
```

	Title	IMDb_rating	MetaCritic	Avg_rating
95	Whiplash	8.5	8.8	8.65
35	Django Unchained	8.4	8.1	8.25
93	Dallas Buyers Club	8.0	8.4	8.20
97	Star Wars: Episode VII - The Force Awakens	8.1	8.1	8.10
3	Arrival	8.0	8.1	8.05
43	Gone Girl	8.1	7.9	8.00
33	The Martian	8.0	8.0	8.00

# Find the Most Popular Trios - I

- Top 5 popular trios

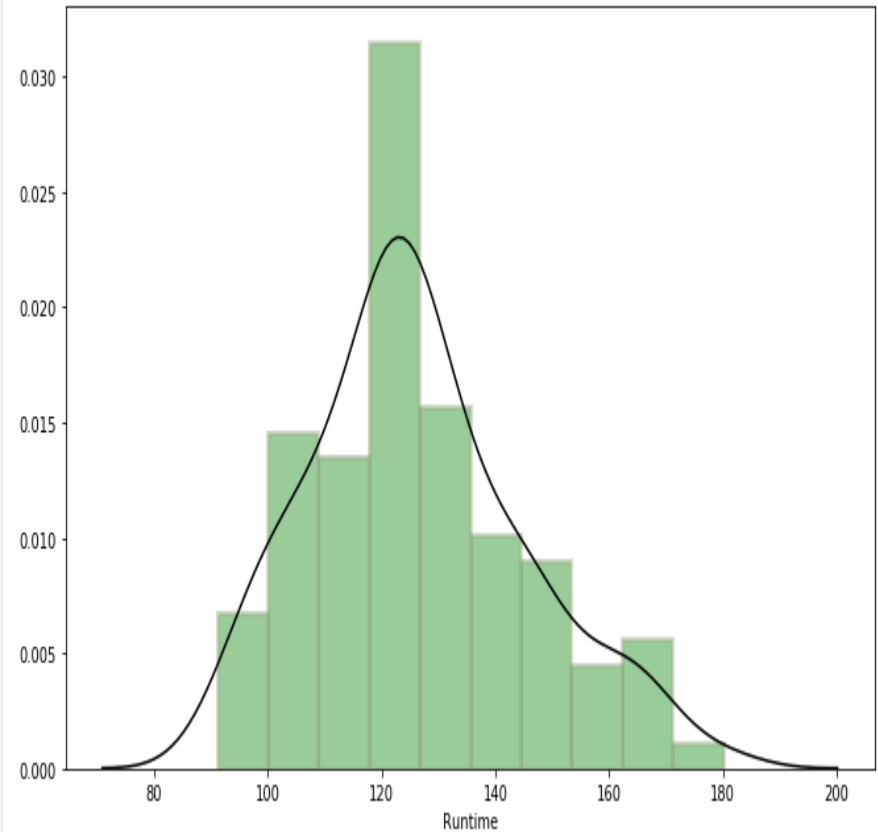
In [21]: group

Out[21]:

	actor_1_name	actor_2_name	actor_3_name	actor_1_facebook_likes	actor_2_facebook_likes	actor_3_facebook_likes	Total likes
0	Dev Patel	Nicole Kidman	Rooney Mara	33000	96000.0	9800.0	138800.0
1	Leonardo DiCaprio	Tom Hardy	Joseph Gordon-Levitt	29000	27000.0	23000.0	79000.0
2	Jennifer Lawrence	Peter Dinklage	Hugh Jackman	34000	22000.0	20000.0	76000.0
3	Casey Affleck	Michelle Williams	Kyle Chandler	518	71000.0	3300.0	74818.0
4	Tom Hardy	Christian Bale	Joseph Gordon-Levitt	27000	23000.0	23000.0	73000.0

# Runtime Analysis

- Plotted a distribution plot to
- find the runtime range most of the movies fall into.
- Most of the movies fall within 120–130 minutes of runtime.



# R-Rated Movies

- Among all the R rated movies that have been voted by the under-18 age group, top 10 movies that have the highest number of votes

In [26]: PopularR

Out [26]:

	Title	CVotesU18
47	Deadpool	4598
36	The Wolf of Wall Street	3622
35	Django Unchained	3250
29	Mad Max: Fury Road	3159
95	Whiplash	2878
31	The Revenant	2619
40	Shutter Island	2321
43	Gone Girl	2286
65	The Grand Budapest Hotel	2083
72	Birdman or (The Unexpected Virtue of Ignorance)	1891

# Combine the Dataframe by Genres

- Created a new dataframe "df\_by\_genre" that contains genre\_1, genre\_2, and genre\_3 and all the columns related to CVotes/Votes.

```
In [41]: genre_top10
```

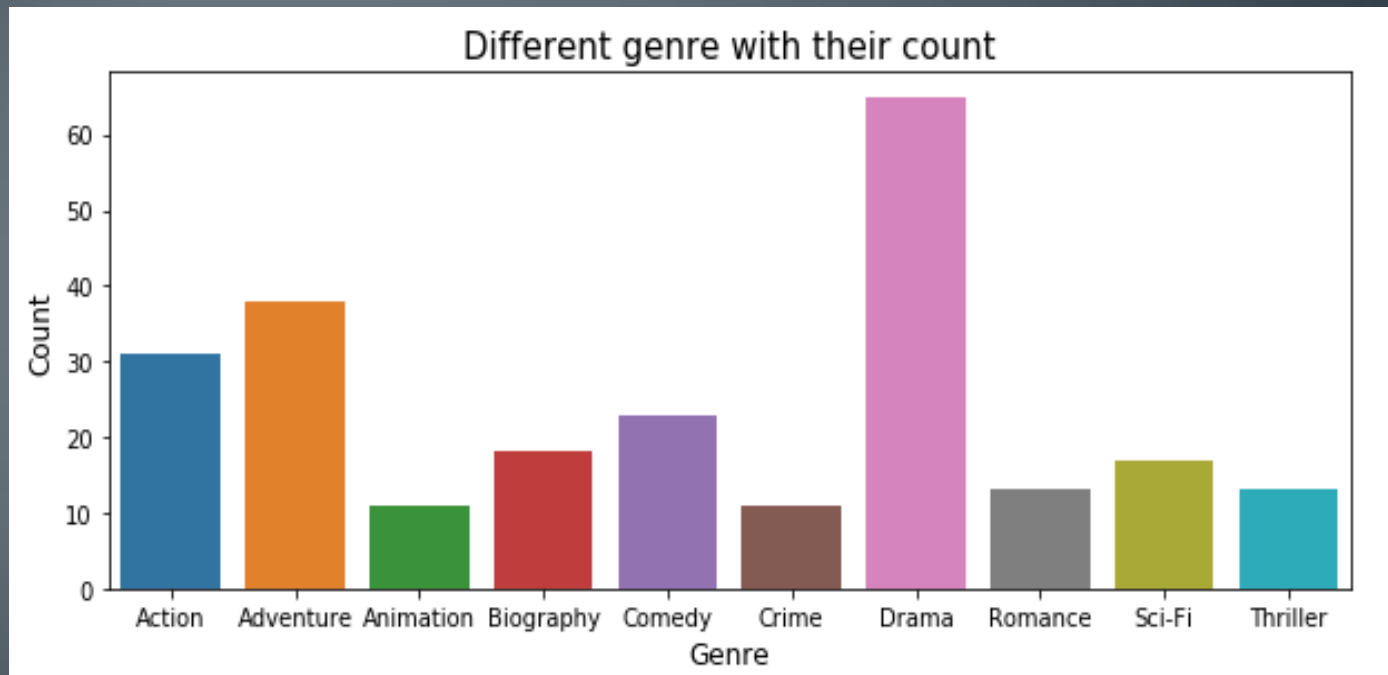
```
Out[41]:
```

	CVotes10	CVotes09	CVotes08	CVotes07	CVotes06	CVotes05	CVotes04	CVotes03	CVotes02	CVotes01	...	Votes3044	Votes3044M	Votes3044F
Action	102144	114433	150895	94262	34688	12693	5386	3064	2115	5524	...	7.74	7.73	7.80
Adventure	94596	105636	138482	86367	31896	11551	4817	2718	1835	4575	...	7.75	7.73	7.87
Animation	61960	72566	104837	65707	22825	7551	2792	1430	911	2290	...	7.76	7.72	7.98
Biography	47333	77867	123948	74054	23644	7702	2984	1639	1145	2849	...	7.73	7.72	7.77
Comedy	60157	77173	108993	69176	26099	9863	4237	2444	1712	3842	...	7.71	7.71	7.75
Crime	52229	87919	129045	74671	25308	8971	3842	2246	1544	3383	...	7.72	7.76	7.61
Drama	52375	75928	109339	66456	23528	8497	3622	2078	1449	3250	...	7.71	7.71	7.72
Romance	42304	53037	82252	54833	21637	8530	3762	2130	1476	3082	...	7.61	7.61	7.66
Sci-Fi	136781	149873	176646	106005	39518	14951	6583	3876	2715	6731	...	7.86	7.85	7.84
Thriller	83207	112730	153336	90446	32003	11534	5021	2918	1982	4433	...	7.74	7.75	7.70



# Genre Counts!

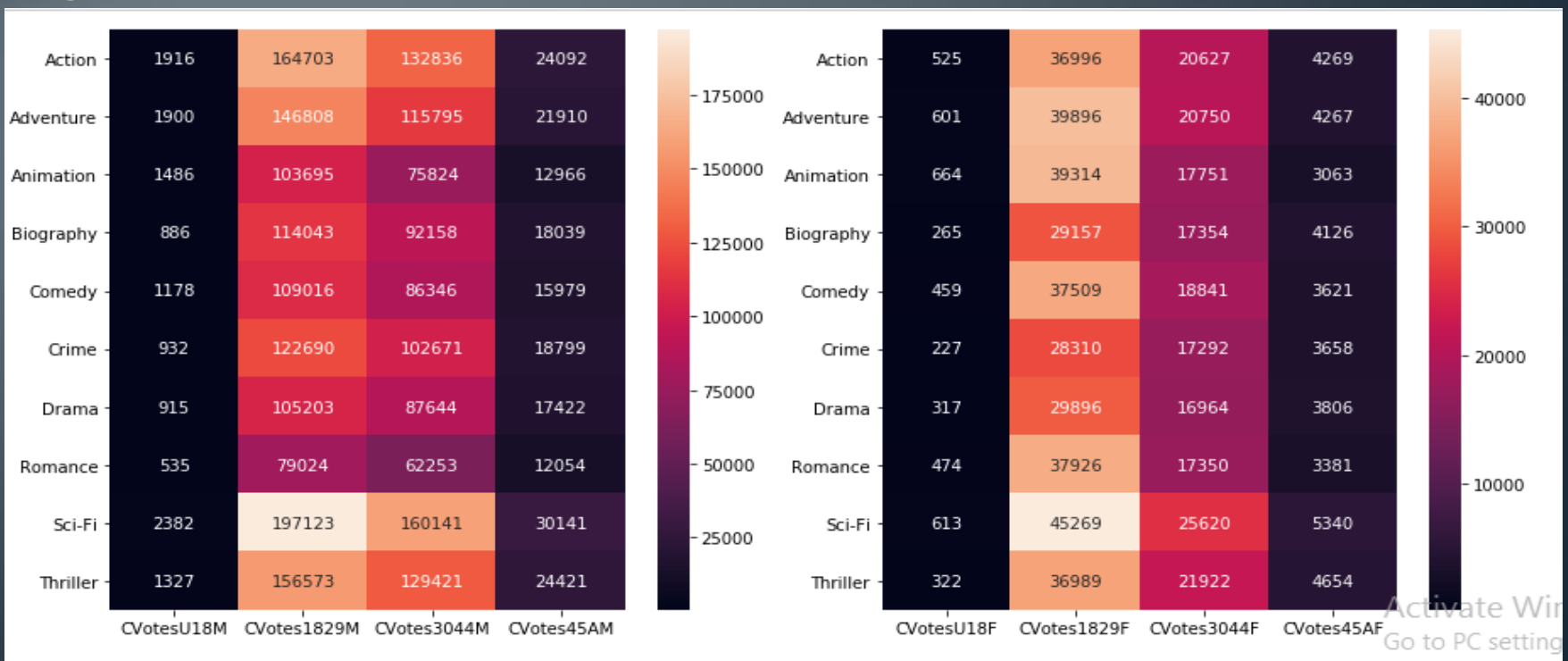
- Created a column called “cnt” which is count and initialized it to one.
- Plotted a bar chart showing different genres vs cnt.



- Drama stands at the top whereas Animation is at the bottom.

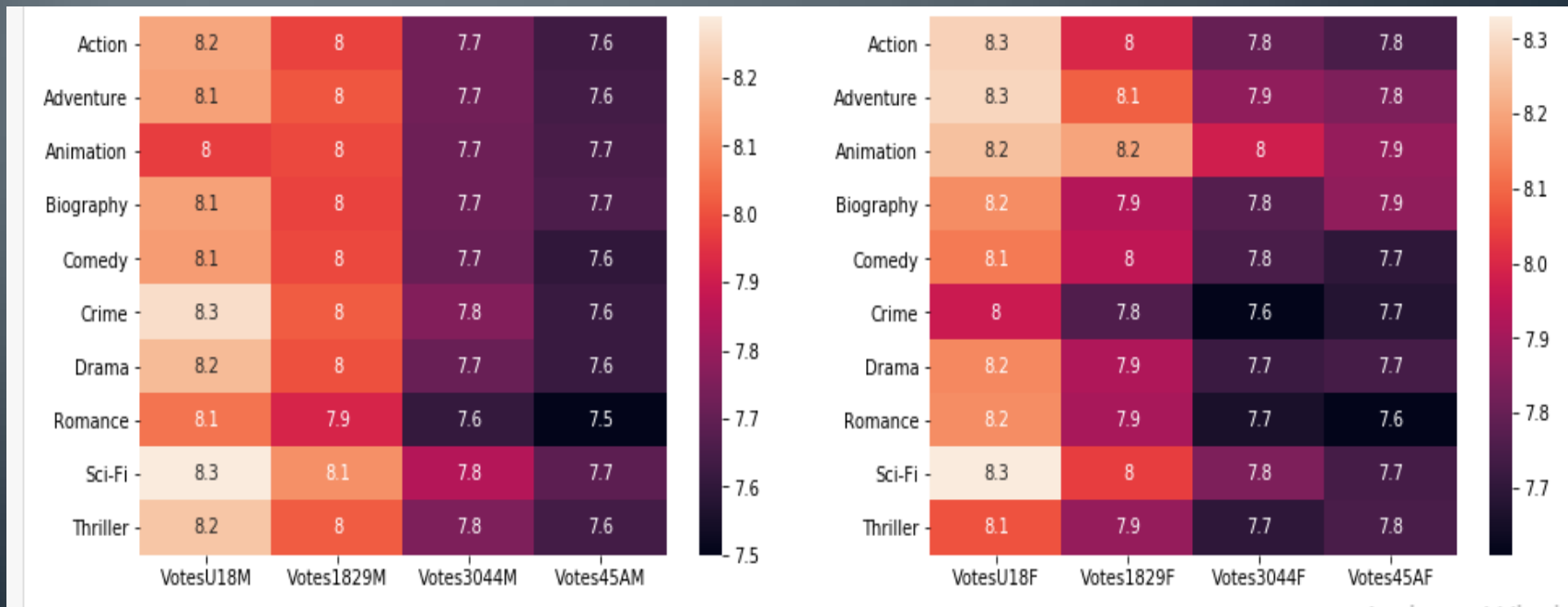
# Gender and Genre

- Plotted the first heat map to show how the average number of votes of males is varying across the genres.
- Age group of 18-29 seems to be the most actively voting for all the genres irrespective of the gender.
- Romance is least voted by Male irrespective of their age compared to other genres.



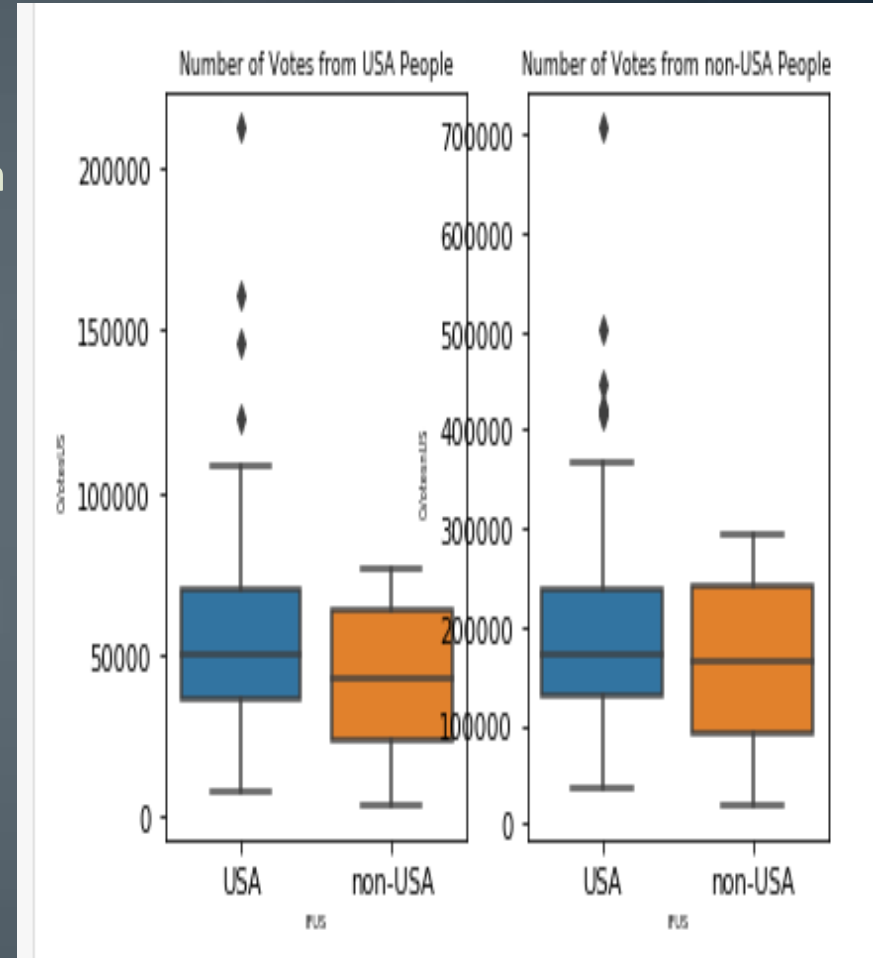
# Gender and Genre

- Plotted the second heat map to show how the average number of votes of females is varying across the genres.
- We can see Animation genres has been voted steadily in Female gender, whereas in the male there is a significant decrease as age increases. It is interesting to observe that females of all age likes animation movies.
- After Sci-Fi, males are more interested in watching crime movies whereas women are interested in action movies.



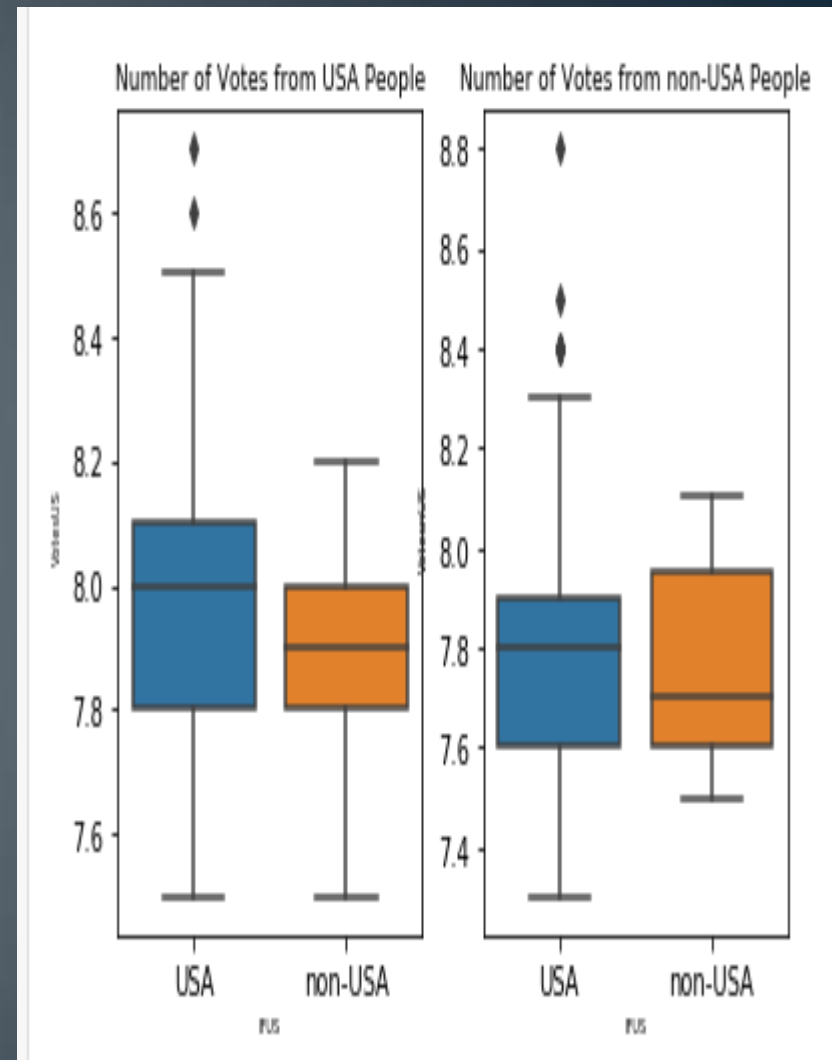
# US vs non-US Cross Analysis

- Created a column "IFUS" which contain the value "USA" if the country of the movie is "USA". For all other countries other than the USA, IFUS should contain the value non-USA.
- Plotted a boxplot that shows the number of votes from the USA people and non-USA people with the help of IFUS column.
- We can observe the number of votes on average is more from non-USA people compared to USA people.
- Overall USA movies got a greater number of votes from both USA and non-USA people compared to non-USA movies.



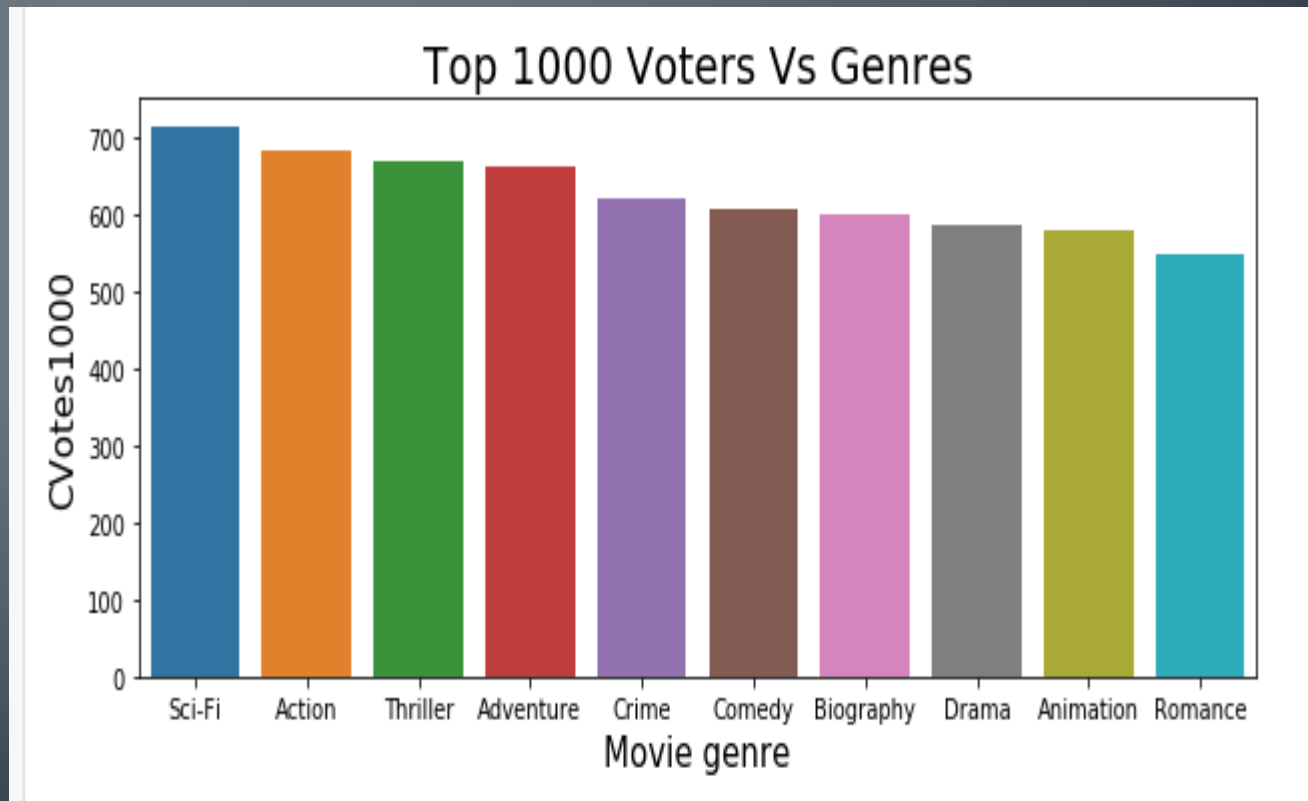
# US vs non-US Cross Analysis

- Plotted a boxplot that shows the ratings from the US people and non-USA people .
- Non-US voters have rated both the US and non-US movies lower as compared to the US voters, which is evident from the medians.
- Some US movies have got exceptionally high ratings from both the US and non-US voters. There are no such extreme ratings for any of the non-US movies.



# Top 1000 Voters Vs Genres

- Plotted a bar chart between the columns CVotes1000 and genre.
- Sci-Fi is the most popular amongst the other genres.
- There is no much significant difference in popularity amongst Action, Thriller and Adventure.



# Recommendations

- The profit earned is almost equal to the budget which means most of the movies have positive impact towards the audience.
- Half of the movies have the average rating equal to or greater than 8.
- As the viewers are much interested in watching movies , the content and the duration of movie should increase because it will help in fetching the profits.
- For all the Cvotes/Votes related, Action seems to be at the first place among the top10 genres.
- Under 18 age group are interested in R-Rated movies as there are highest number of votes from that age group.
- Drama seems to be the most popular genre as it has the highest number of viewers compared to all the other genres.
- For Cvotes related, 18-29 age group seems to be most actively voting for any genre irrespective of the gender.
- When compared with US and non-USA movies, USA movies got the highest number of votes and ratings as it is clear from the boxplot.
- From the data collected for the Top 1000 voters, Sci-Fi seems to be the most viewed genre and the director should focus more on this as it will fetch them profit.