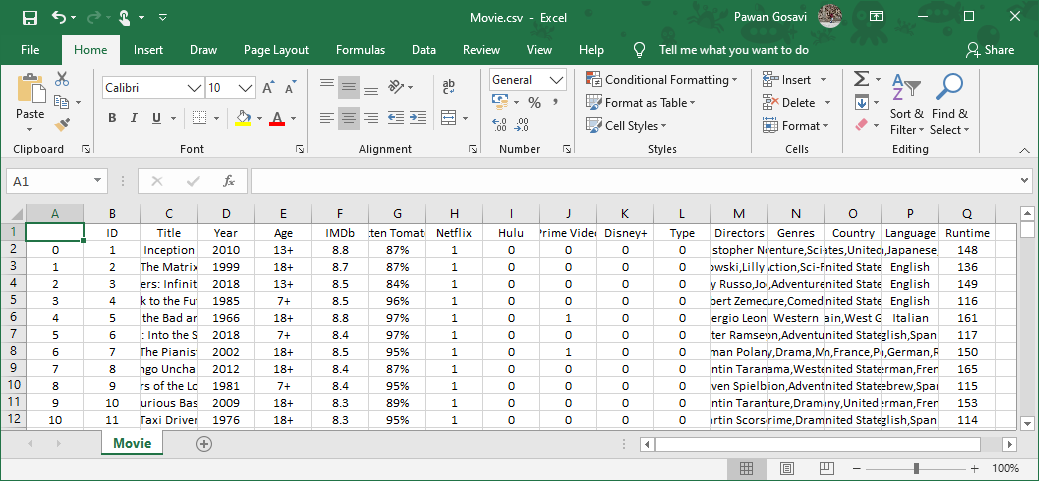
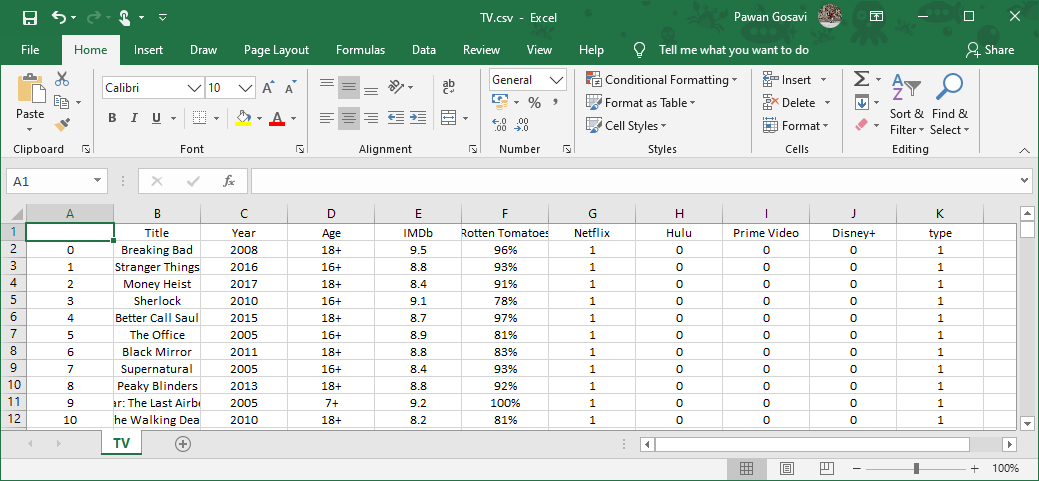
**Data Sets**

**Movie.csv**



**TV.csv**



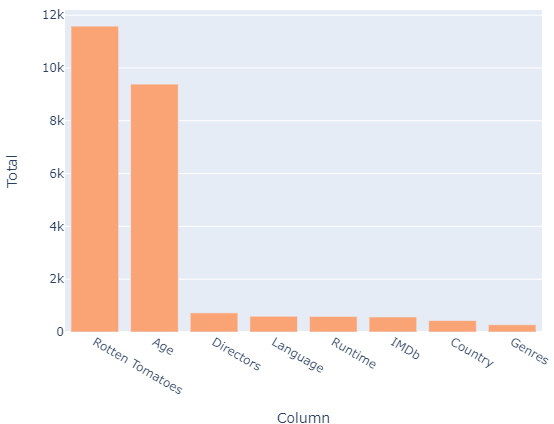
**First Key Observations:**

* Oldest movie is from 1902.
* There is a total of 16744 movies in this dataset.
* Average runtime of all the movies is about 93 minutes.
* There is an entry with runtime as 1256 minutes. This looks odd. We need to verify this data.
* More than 50% of the data is missing for Rotten Tomatoes ratings and Age restriction columns.

**Percentage of missing values**

|  |  |  |
| --- | --- | --- |
|  | column name | Percentage missing values |
| 6 | Rotten Tomatoes | 69.19 |
| 4 | Age | 56.08 |
| 12 | Directors | 4.34 |
| 15 | Language | 3.58 |
| 16 | Runtime | 3.54 |
| 5 | IMDb | 3.41 |
| 14 | Country | 2.60 |
| 13 | Genres | 1.64 |
| 10 | Disney+ | 0.00 |
| 11 | Type | 0.00 |
| 0 | Unnamed: 0 | 0.00 |
| 9 | Prime Video | 0.00 |
| 1 | ID | 0.00 |
| 7 | Netflix | 0.00 |
| 3 | Year | 0.00 |
| 2 | Title | 0.00 |
| 8 | Hulu | 0.00 |

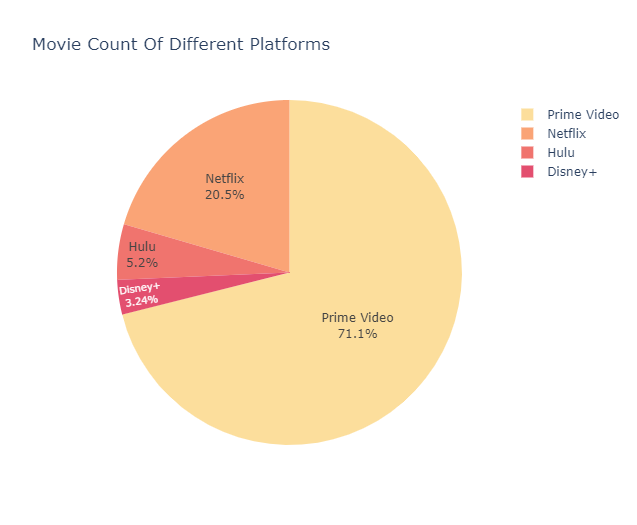
**MISSING VALUES**



**PLATFORMS**

There are 4 Platforms for this Data.





Prime Video has the majority of movies on its platform.

**TITLE**

3560 titles are streaming in Netflix, 903 titles are streaming in Hulu, 12354 titles are streaming on Prime video, 564 titles are streaming on Disney+.

We have close to 12k titles which are streaming on Prime video whereas 3.5k titles comes from Netflix. Hulu and Disney+ titles are very less compared to the two.

Are there any titles available in all the 4 streaming platforms?

Well, no movie title is present in all the 4 streaming platforms. Close to 96 % of the movie titles available in this dataset are present in only one of the 4 streaming platforms. Less than 4 % of the titles are available in 2 platforms whereas it is very rare to see a title available in 3 platforms. Let’s do a quick check on the number of titles and other details for the titles present in 3 streaming platforms.

There are 10 movie titles in all present in 3 streaming platforms.

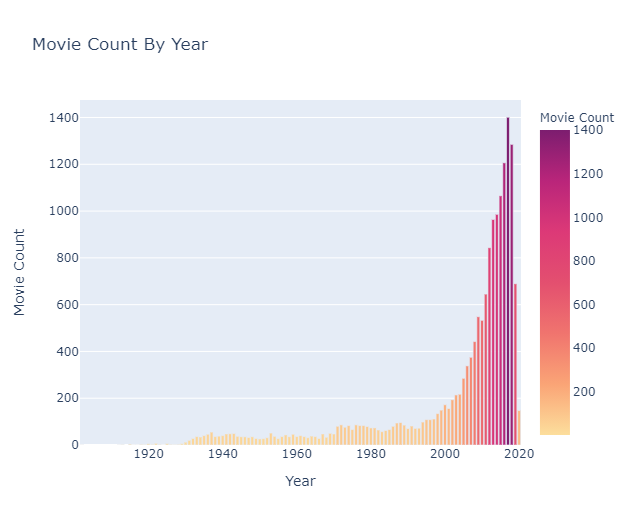


Netflix, Hulu and Prime video seem to have most common titles among the 4 platforms. There is a great overlap between Hulu and Prime video here with each differing only with 1 title. And that two titles - Zapped and Amy seem to be available for streaming in Netflix.

More Hulu titles are available either in Netflix and Prime video compared to the Disney+.

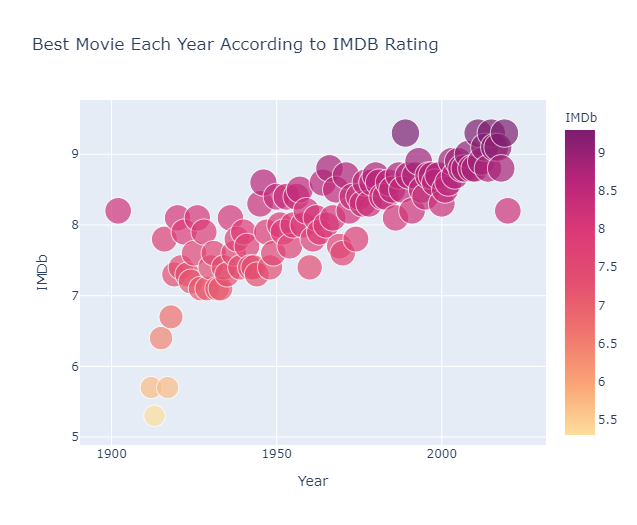
**YEAR**

**Movie Count by Year**



Movies has been increasing year by year and the highest movies were released in the year 2017 in which 763 were released on Prime Video and 569 on Netflix.

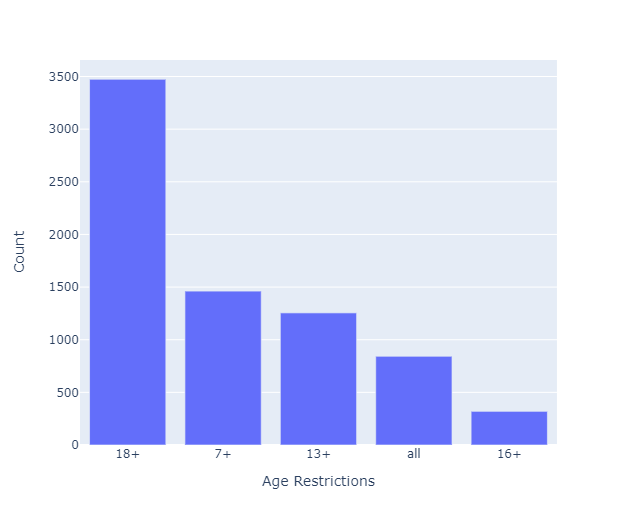
**Best Movie Each Year According to IMDB Rating**

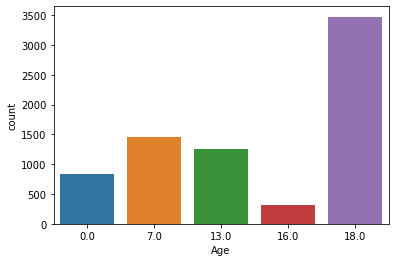


Best Movies in the last few years have been the best in all the years in terms of IMDB Rating.

**AGE**

**Age Restrictions**





0 indicates no restriction in the age group in the above dataset

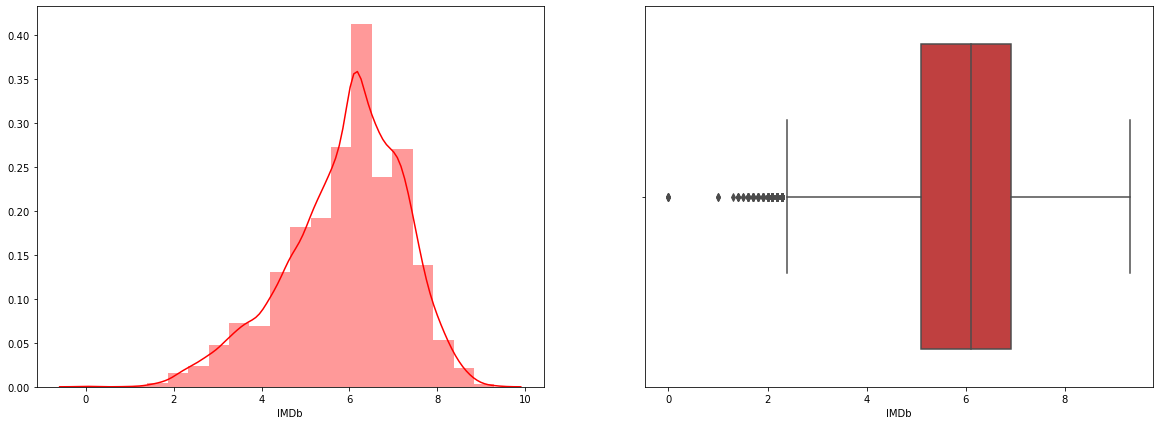
it is clearly visible that most of the shows in the OTT platforms are targeted for age group greater than 18.

Followed by age group of 7 and 13 which may include movies of kid genre.

**IMDb**

**IMDb Ratings**

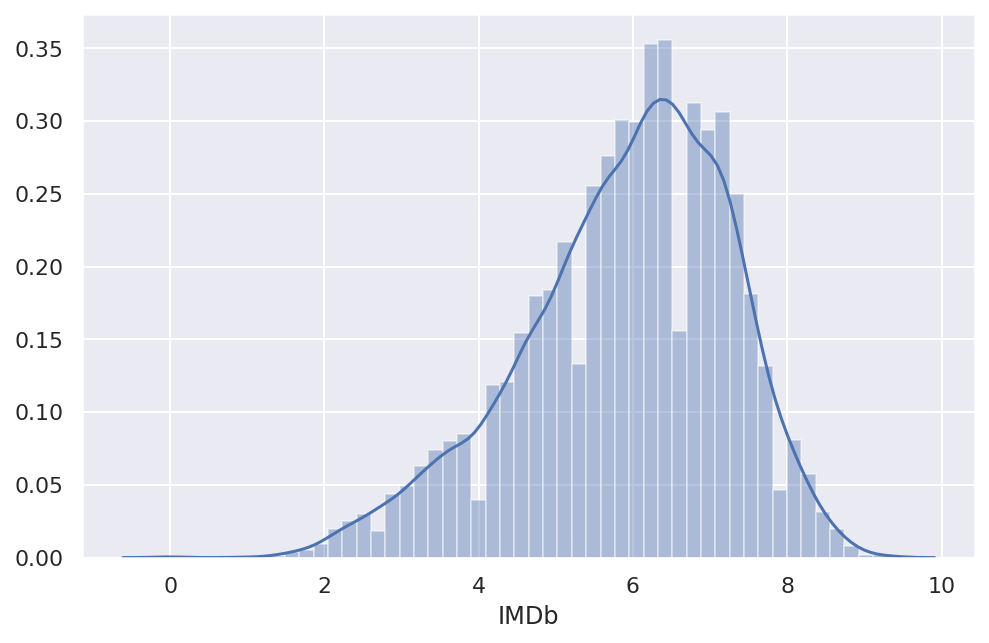
A high IMDb rating makes an insignificant contribution towards a movie's success. Rating says how good the movie is and decides whether it can be watched or not



The average rating lies between 5-7. Also, there are some outliers present.

The movie is considered to be extraordinary when the rating lies above 8. The movie is considered to be good if the rating lies between 6 - 8.

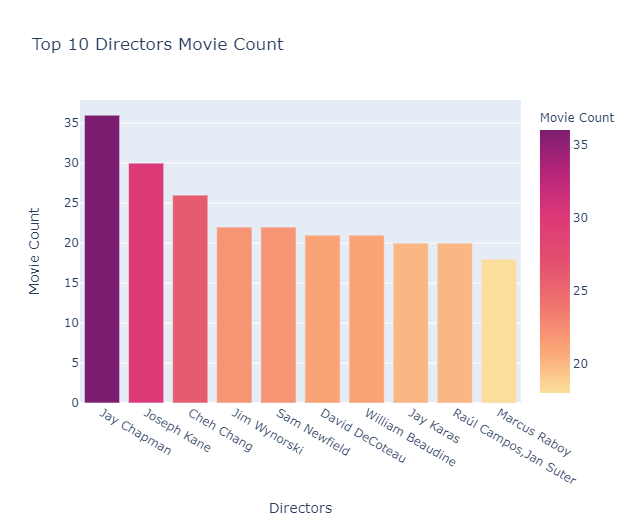
**IMDb Ratings Distribution**



IMDb ratings of the movies roughly follows normal distribution with mean of about 6.5.

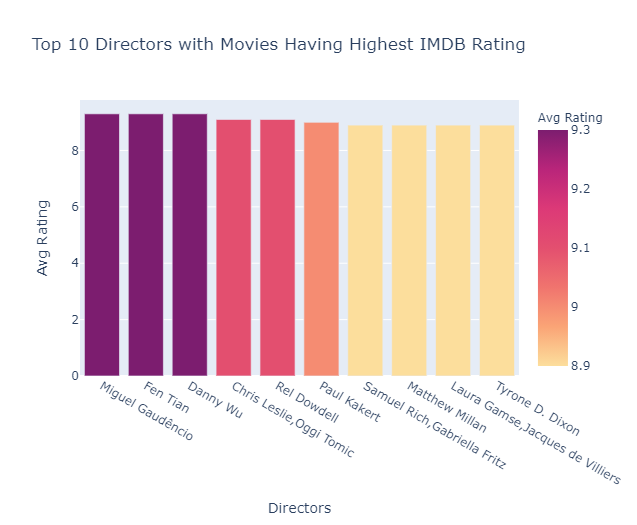
**DIRECTORS**

**Top 10 Directors Movie Count**



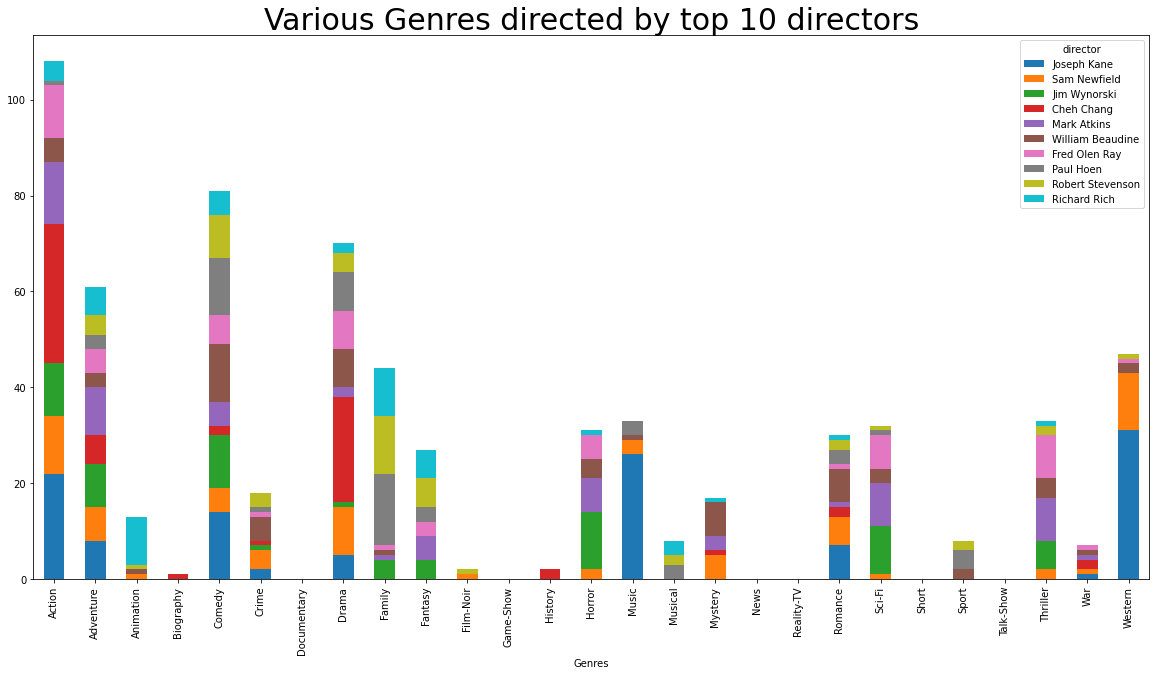
Jay Chapman has the directed the greatest number of movies and most of them are present on Prime Video.

**Top 10 Directors with Movies Having Highest IMDB Rating**



The directors who made highest ranked IMDB movies have only one movie count!

**Genres Directed by Top 10 Directors**



Based on the above visualization we can infer that most of the action, comedy, music and western genre movies are directed by Joseph Kane.

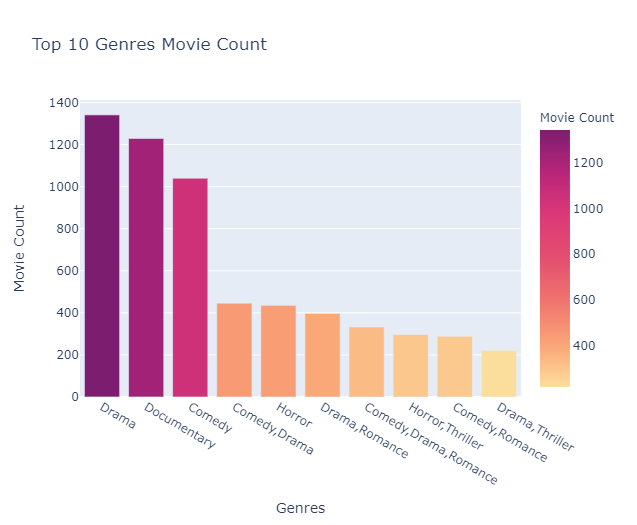
Also, a wide range of action and drama genre movies are directed by Cheh Chang.

If you want to go with thriller movies, then you'll have a wide availability of movies directed by Fred Olen Ray and Mark Atkins.

William Beaudine has directed movies in almost all the above-mentioned genres.

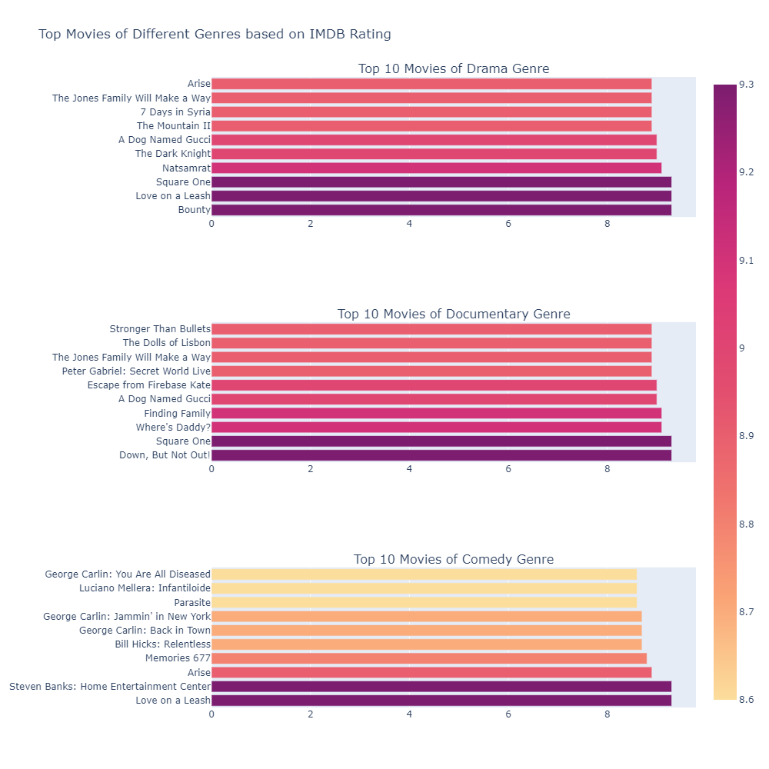
**GENRES**

**Top 10 Genres Movie Count**



Drama, Documentary and Comedy are the most famous genres among all.

**Top Movies of Different Genres based on IMDB Rating**



Drama, Documentary and Comedy some Movies with 9+ Ratings.

**USING KMEANS CLUSTERING TO PLOT DIFFERENT GENRES**

**K-means clustering** is one of the simplest and popular unsupervised machine learning algorithms.

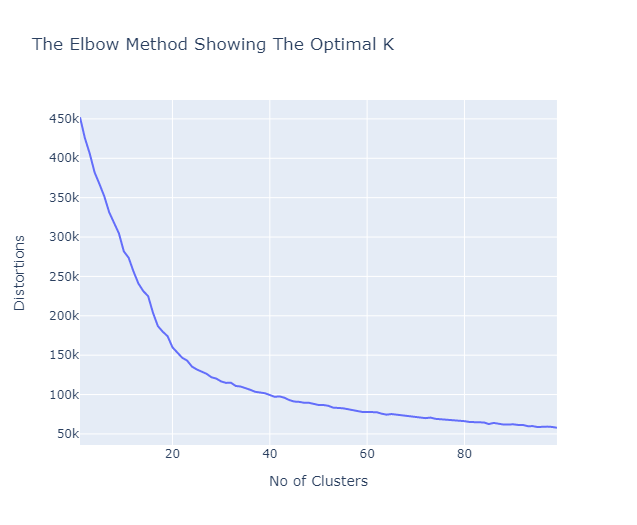
A cluster refers to a collection of data points aggregated together because of certain similarities. You’ll define a target number k, which refers to the number of centroids you need in the dataset. A centroid is the imaginary or real location representing the center of the cluster. Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares. In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

**StandardScaler** - It standardizes a feature by subtracting the mean and then scaling to unit variance. Unit variance means dividing all the values by the standard deviation.

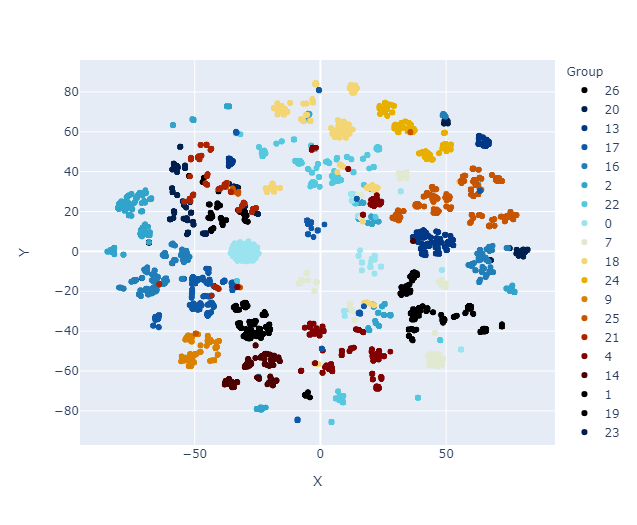
**TSNE** - t-Distributed Stochastic Neighbour Embedding (t-SNE) is a technique for dimensionality reduction that is particularly well suited for the visualization of high-dimensional datasets.

**Elbow Method**

The elbow method helps to choose the optimum value of ‘k’ (number of clusters) by fitting the model with a range of values of ‘k’. Here we would be using a 2-dimensional data set but the elbow method holds for any multivariate data set.



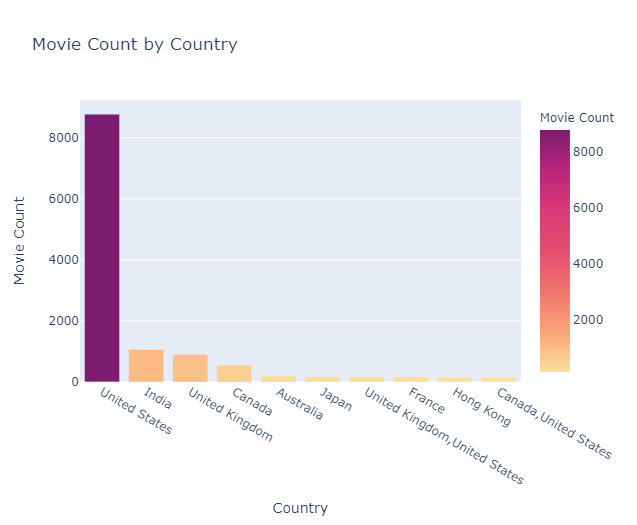
We can observe that the “elbow” is the number 27 which is optimal for this case. We can also verify this because there are 27 different genres so this result was pretty much expected.



Great! We can even check that clustering is pretty accurate by checking the genres of movie with its nearest neighbour.

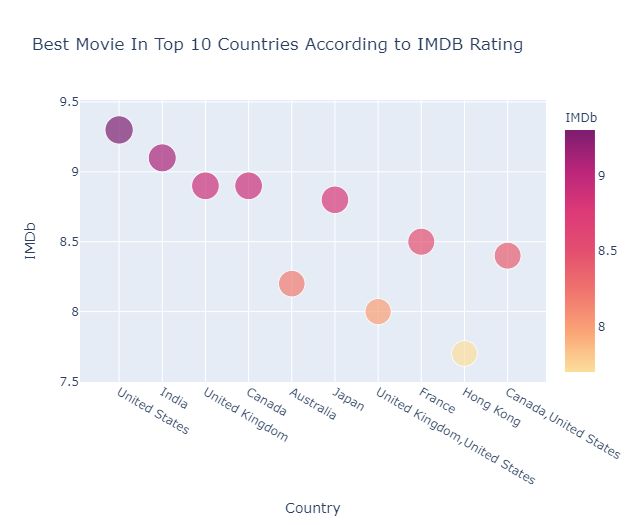
**COUNTRY**

**Movies Count by Country**



Majority of the movies are of United States with 6817 movies on Prime Video and 1305 movies on Netflix.

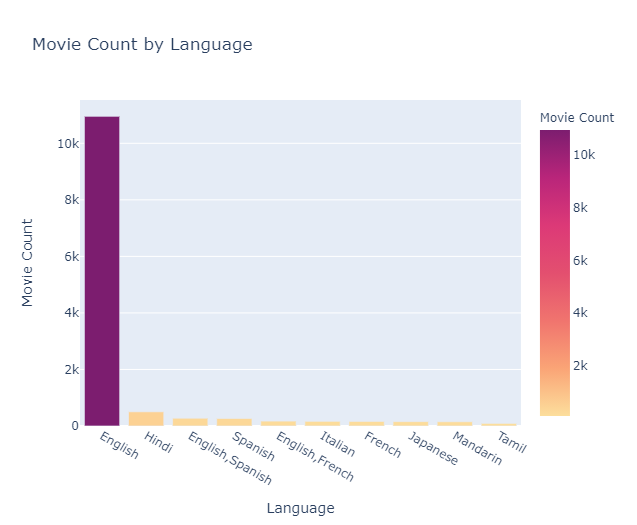
**Best Movie in Top 10 Countries According to IMDB Rating**



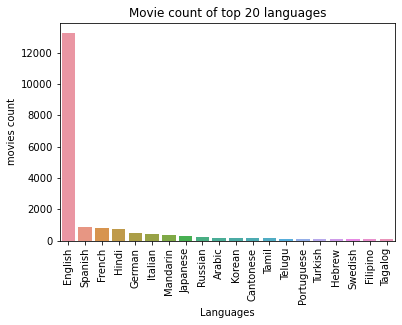
Since majority of the movies are of United States, it was expected that the best movies from United States will also be best among others.

**LANGUAGE**

**Movie Count by Language**

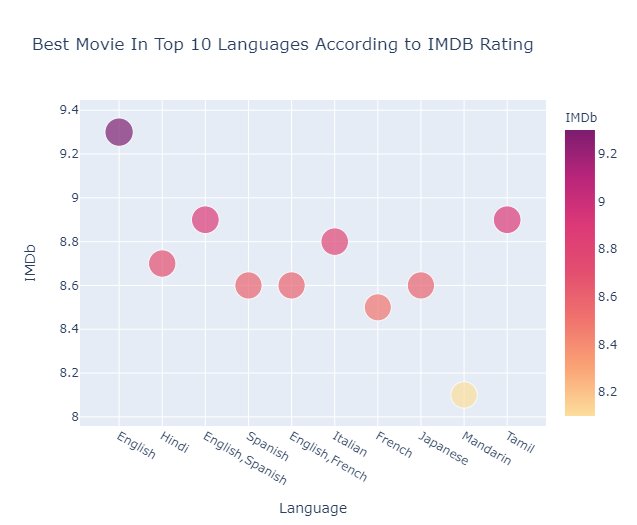


Almost two-third of the movies in this dataset are of English language. English is the most common of all languages. There are 8617 English movies on Prime Video and 1624 on Netflix. It is clearly visible that most of the movies are from English which shows that all the platforms are trying to get viewership across the globe.



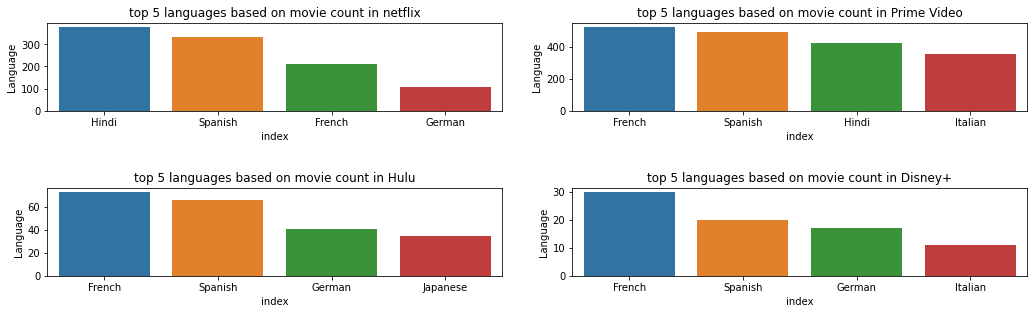
Indian regional languages like Hindi, Tamil, Telegu, Punjabi is among the top 20 which shows that all the OTT platforms has a good reach and viewership in India. A wide range of European language can be seen in the top 20 which again shows the popularity among the European nations.

**Best Movie in Top 10 Languages According to IMDB Rating**



Since majority of the movies were in English, it was expected that the best movie from English will also be best among others.

**Top 5 languages based on movie count across the OTT Platforms**

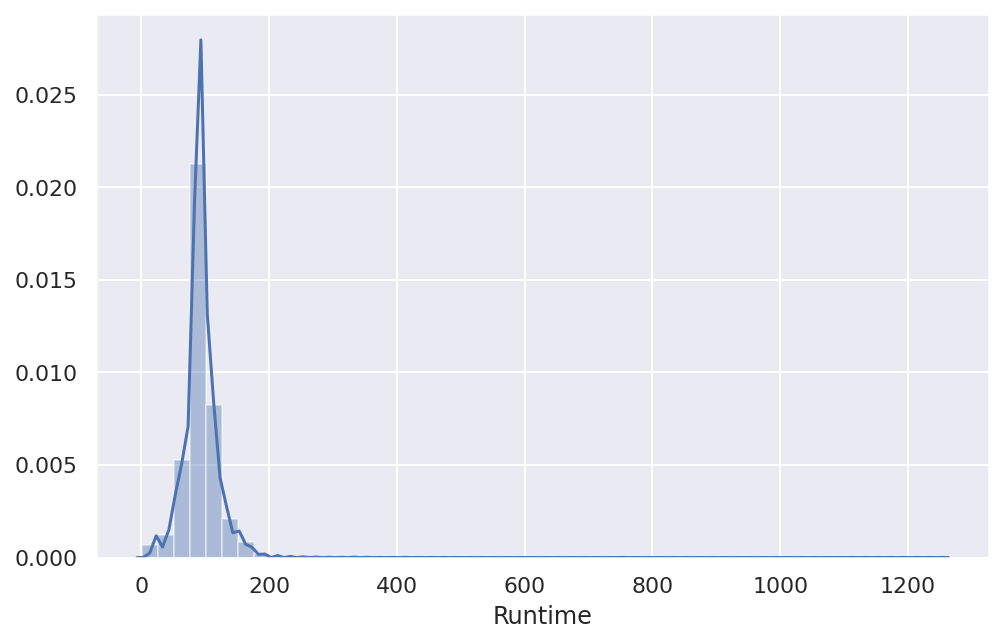


The above visualization depicts the top 5 languages (other than English) having the highest movie count in different OTT platforms.

As seen earlier all OTT Platforms are having higher number of English language movies. If you want to watch Hindi movies you can go with Netflix and Prime Video with more than 400 movies.

Spanish language movies are equally spread and having more count across all the OTT platforms. German movies can be viewed in Netflix with more than 100 movies.

**RUNTIME**



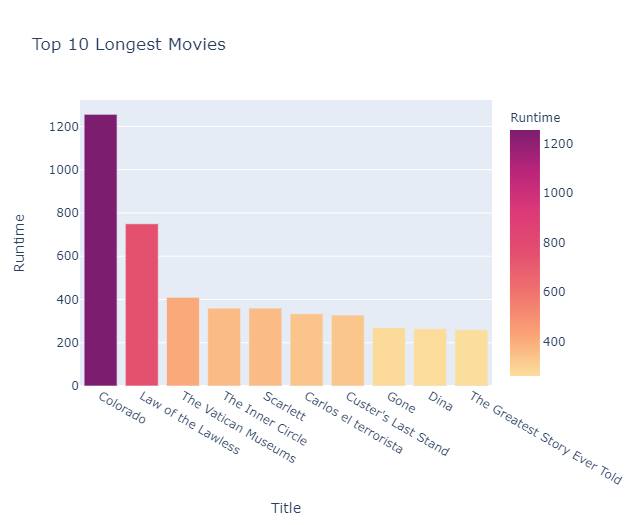
Most of the movies have about 90-100 minutes of runtime. Runtime is positively skewed since there are some movies with very large runtime

**Average Runtime on different Platforms**



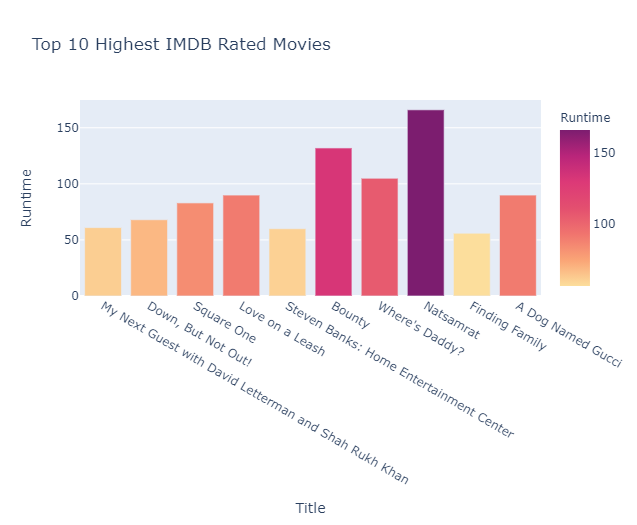
Average Runtime of movies on all platforms is similar.

**Top 10 Longest Movies**



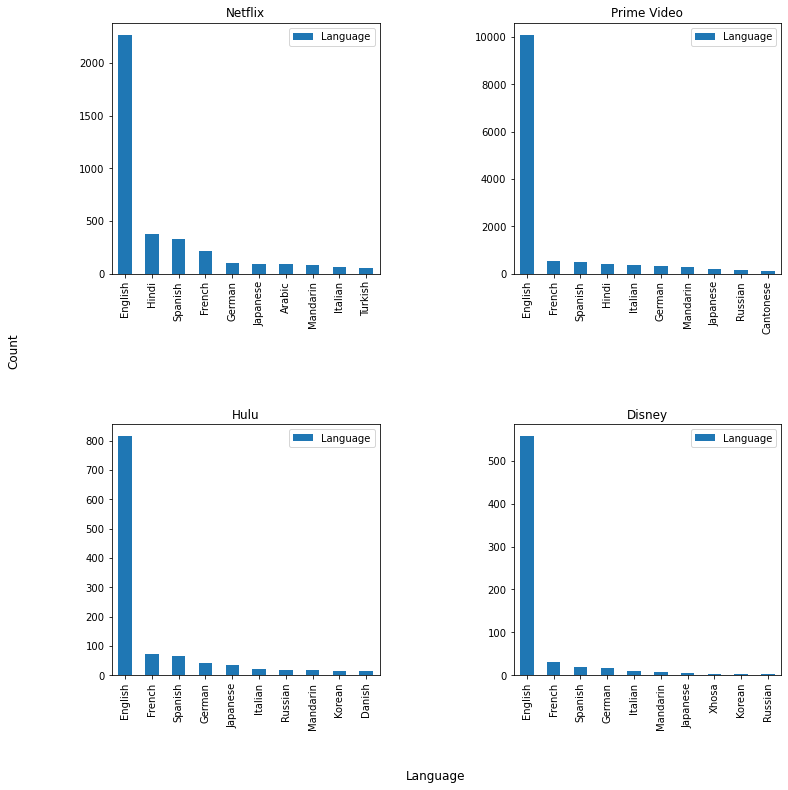
Colorado has the highest runtime of whopping 1256 minutes and is only present on Prime Video.

**Top 10 Highest IMDB Rated Movies**



Of all the highest rated movies, Natsamrat has the highest runtime of 166 minutes and is present only on Netflix.

**Language Vs Platforms**



All the platforms have a greater number of movies on 'English' language.

Netflix: Followed by 'English’, Hindi and Spanish movies comes in line.

Prime Video: Followed by 'English', French and Spanish movies comes in line. 'Hindi' movies stand behind these three.

Hulu: Followed by 'English’, French and Spanish movies comes in line. 'German' movies stand behind these three.

Disney+: Followed by 'English’, French and Spanish movies comes in line. 'German' comes after all these.

**Language Vs Ratings**



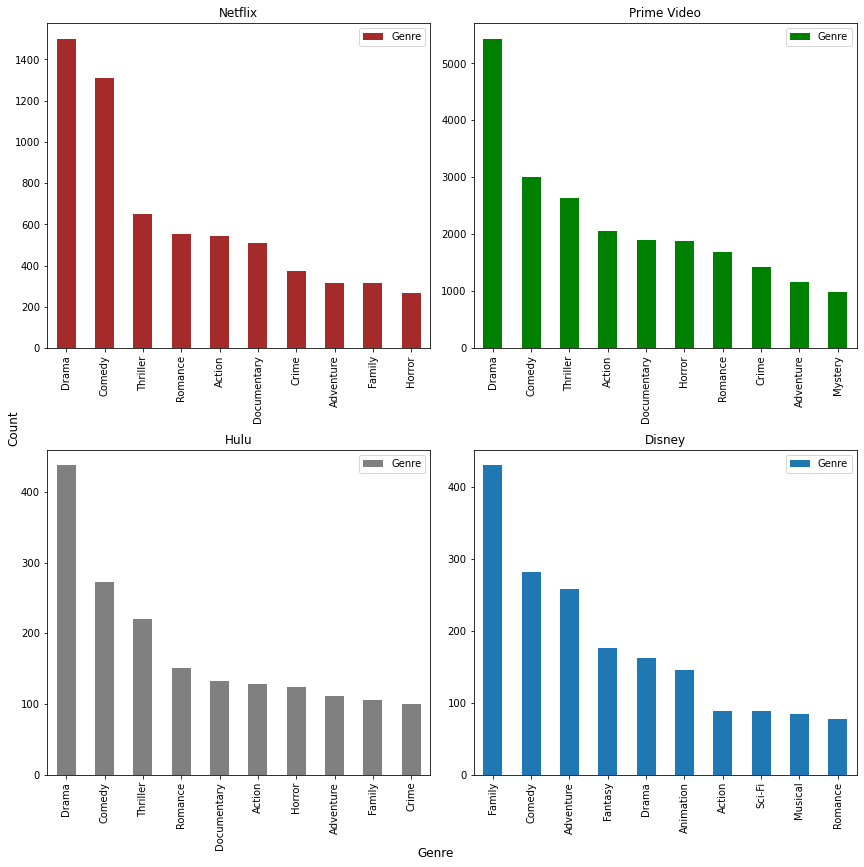


Top 10 languages have been chosen based on the distribution of movies and average ratings have been calculated.

Since maximum movies are of language 'English', median rating of the movies of 'English' is calculated.

It says that most of the 'English' movies has a median rating of '6.0'

**Genre Vs Platforms**



Netflix: Genre of 'Drama' has 1500+ movies. It is followed by comedy which is around 1300+. Then 'Thriller', 'Romance', 'Action', 'Documentary' has equally distributed.

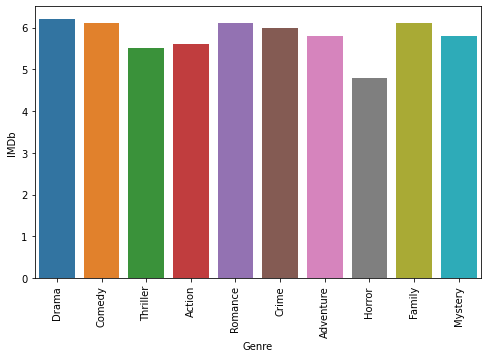
Prime Video: Genre of 'Drama' has 5000+ movies. It is followed by comedy and thriller which is around 2500+ movies. 'Action' comes next.

Hulu: Here also 'Drama' comes first which has around 450+ movies. 'Comedy' and 'Thriller' follows 'Drama'.

Disney: Unlike other 3 platforms, 'Family' has the leading position with 450+ movies. Then 'comedy' and 'adventure' comes into play followed by 'Fantasy'.

From this we infer that, 'Disney+' is more suitable for kids as well as family. Netflix and Prime Video is more suitable for people who enjoy 'Drama' and 'Comedy'.

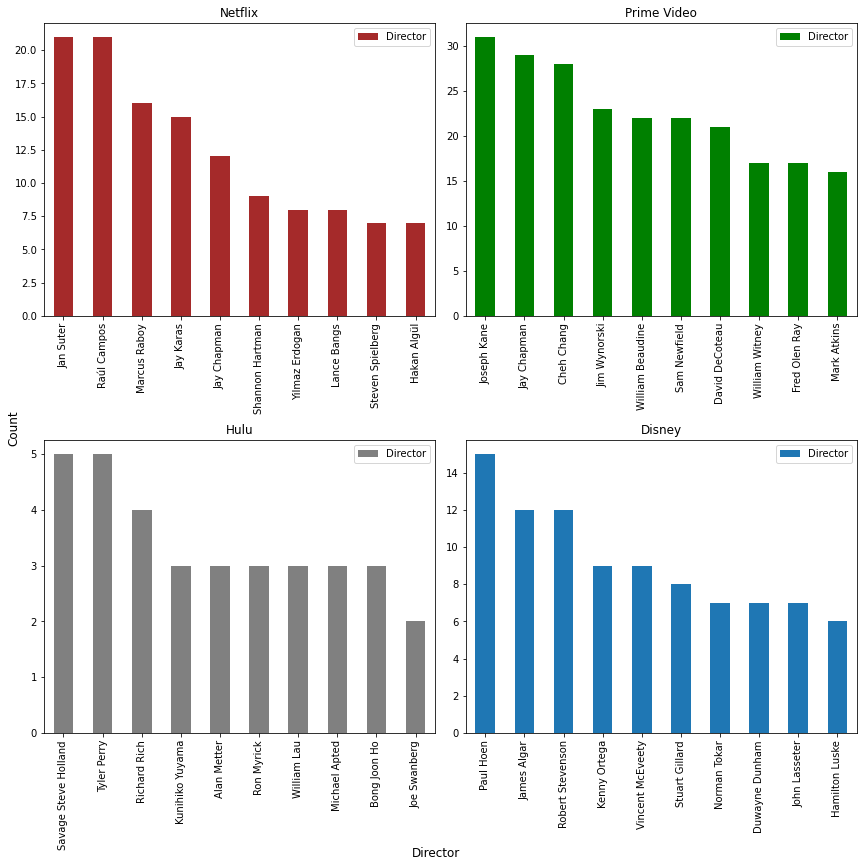
**Genre Vs Ratings**



Based on the movie count, Top 10 genres have been chosen.

With help of this, we can say that all genres have the median rating around 5.8 - 6.2. 'Family' and 'Drama' has highest median rating of 6.2.

**Director Vs Platforms**



Relationship between Director and platforms is shown.

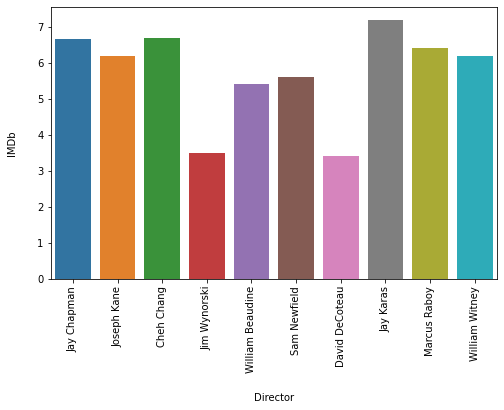
Netflix: 'Jan Suter' and 'Raul Campos' leads the Netflix platform by their movies with 20+ movies each.

Prime Video: 'Joseph Kane' leads the prime video platform with more than 30+ movies. Then followed by 'Jay Chapman'.

Hulu: 'Tyler Perry' and 'Steve Holland' has around 5+ movies in Hulu.

Disney+: 'Paul Hoen' leads the Disney platform by 15+ movies. 'James Alagar' follows him with 11+ Movies.

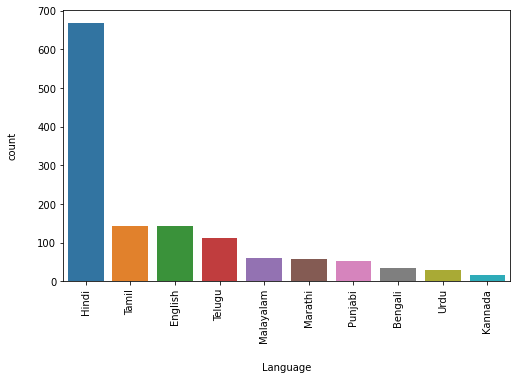
**Director Vs Ratings:**



Based on the total movie count, top 10 directors have been chosen and their median ratings have been displayed. 'Jay Karas' grabs the first position based on the ratings followed by 'jay Chapman'.

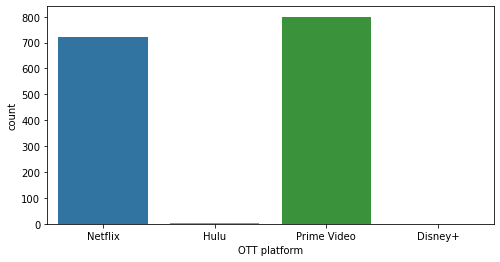
**Movies in India**

**Indian Language Movies**



When it comes to Indian movies, 'Hindi' hits every other language having movie count of 650+. Then comes 'Tamil' and 'English' with count of 150+ movies each.

**Indian Movies vs OTT platforms**



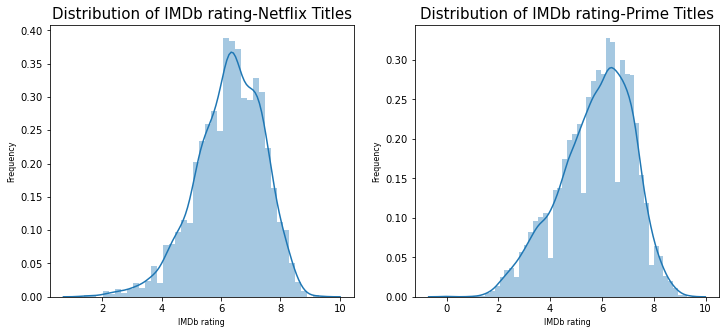
Prime video and Netflix are almost equal when comes to Indian movies. Prime video has 800+ and Netflix has 700+ movies. These two platforms are preferred for Indian movies particularly 'Hindi' Language.

**Netflix Vs Prime Movies**

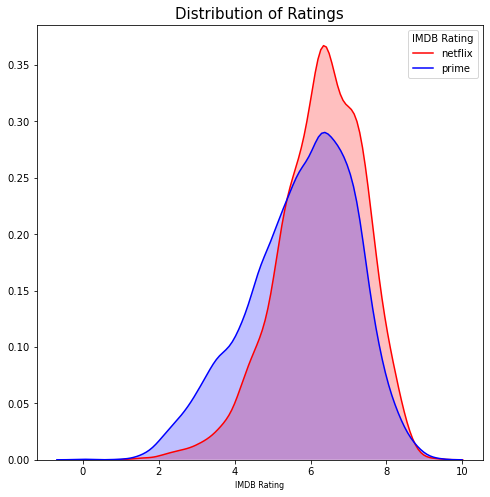
From the previous analysis, we found that there are 3560 titles available in Netflix & 12354 titles are available in Prime. Let’s make a comparison between these two streaming platforms and infer our observations.

Analysing the age restrictions for the titles, it is seen that both the streaming platforms have 20% of their titles in 18+ rating.

The percentage of 13+ rated movies is higher in Netflix compared to Prime. While there is a 4% difference between the 13+ titles there is a 2% difference in 7+ ratings.

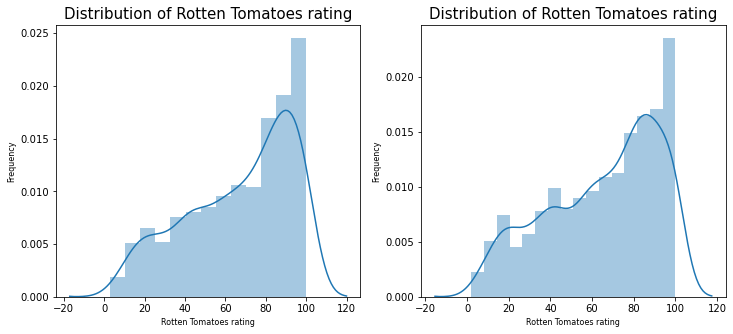


There is a difference in the IMDB ratings between Netflix and Prime.

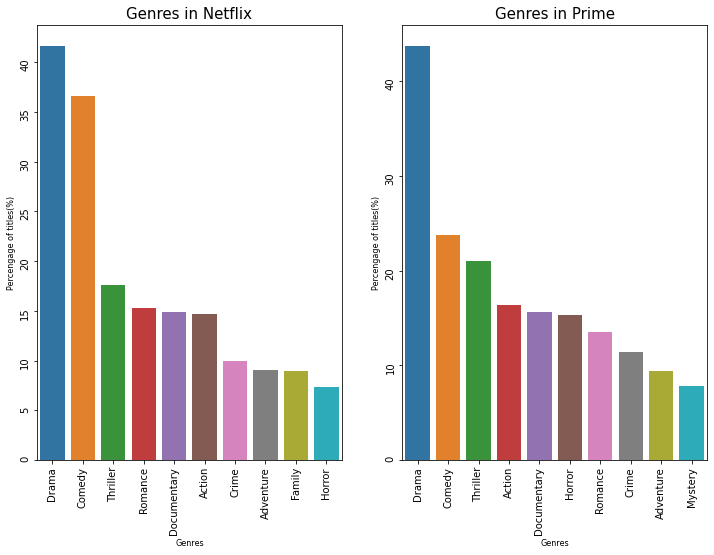


The distribution is close to bell shaped curve (not a perfect) with the peak between 6-7 for Netflix and 5-6 for Prime.

There are few titles having rating greater than 9. Compared to Netflix, there are more titles with lesser IMDB ratings (between 2-5 range) available in Prime whereas in the higher ratings scale, Netflix seem to be dominant.



Ratings in rotten tomatoes have been on the higher side with titles available in the range between 80-100 for both Netflix and Prime. Though there is a difference between the two streaming platforms when it comes to IMDB ratings, similar ratings are seen for the titles available in Prime and Netflix.



The top 3 genres available in Netflix and Prime are the same.

Though the difference in percentage is relatively large for the first three genres, it is less for the subsequent genres. They seem to be available in equal percentages.

Though the percentage of drama titles are very close between Netflix and Prime, more titles in Comedy genre are available in Netflix compared to Prime whereas more titles in Thriller genre are available to stream in Prime compared to Netflix.

While 16% of the movie titles belong to Romance genre in Netflix, an equal percentage of titles belonging to Action genre is available in Prime.

Mystery genre of movies seem to occupy the 10th place whereas the genre is not in top 10 for Netflix.

So, when it comes to genres, both Netflix and Prime have almost equal representation from Drama, Comedy and Thriller.

**Analysis & Observations**

(There were a lot of missing points in the data set. An analysis has been done excluding those point)

1. 96 % of the titles listed in this dataset are available to view in only one platform. Therefore, each platform has got his own unique catalogue of movies and genres to pull its customers.
2. Netflix has got more movies with higher IMDB rating of 6-7 when compared to Prime. This can be noticed from the histogram plot of genres between Netflix and prime.
3. When it comes to ratings on rotten tomato, both the platforms have got similar ratings for the titles.
4. When it comes to genres, Drama, Comedy and Thriller are the most available genres in both Netflix and Prime.
5. As per the available data, the "Prime video" has the highest number movies followed by "Netflix". "Disney+" has the lowest number of movies casted.
6. Most of the movies produced were "Drama" followed by "Comedy". Almost negligible number of movies produced were of "Game-Show" genre.
7. The number of movies released under the genres "Reality-TV", "Talk-Show", "Film-Noire" and "News" were a lot less as compared to other genres.
8. The number of movies produced throughout under "Musical" and "Short" genres were almost similar.
9. As there were a large number of missing data, after removing those values it is seen that most of the movies produced had "Rotten Tomatoes" rating >81% followed by 61% - 80%. The least number of movies produced had <20%.
10. Most of the movies released were of age group 18+. The least number of movies produced were of age group 16+.
11. Majority of the movies were released between 2000 and 2020. The least number of movies released were between 1900 and 1900.
12. According to the data, it is seen that the frequency of movie production increased along with an increase in Year. They both exhibit a positive correlation.
13. Majority of the movies has an IMDb rating ranging between 5-6 followed by 7-8.
14. The least number of movies got an IMDb rating between 9-10.
15. On the basis of given data, we can say that majority of the movies in these 4 OTT platforms are average or more than average IMDb rated.