

Table of Contents

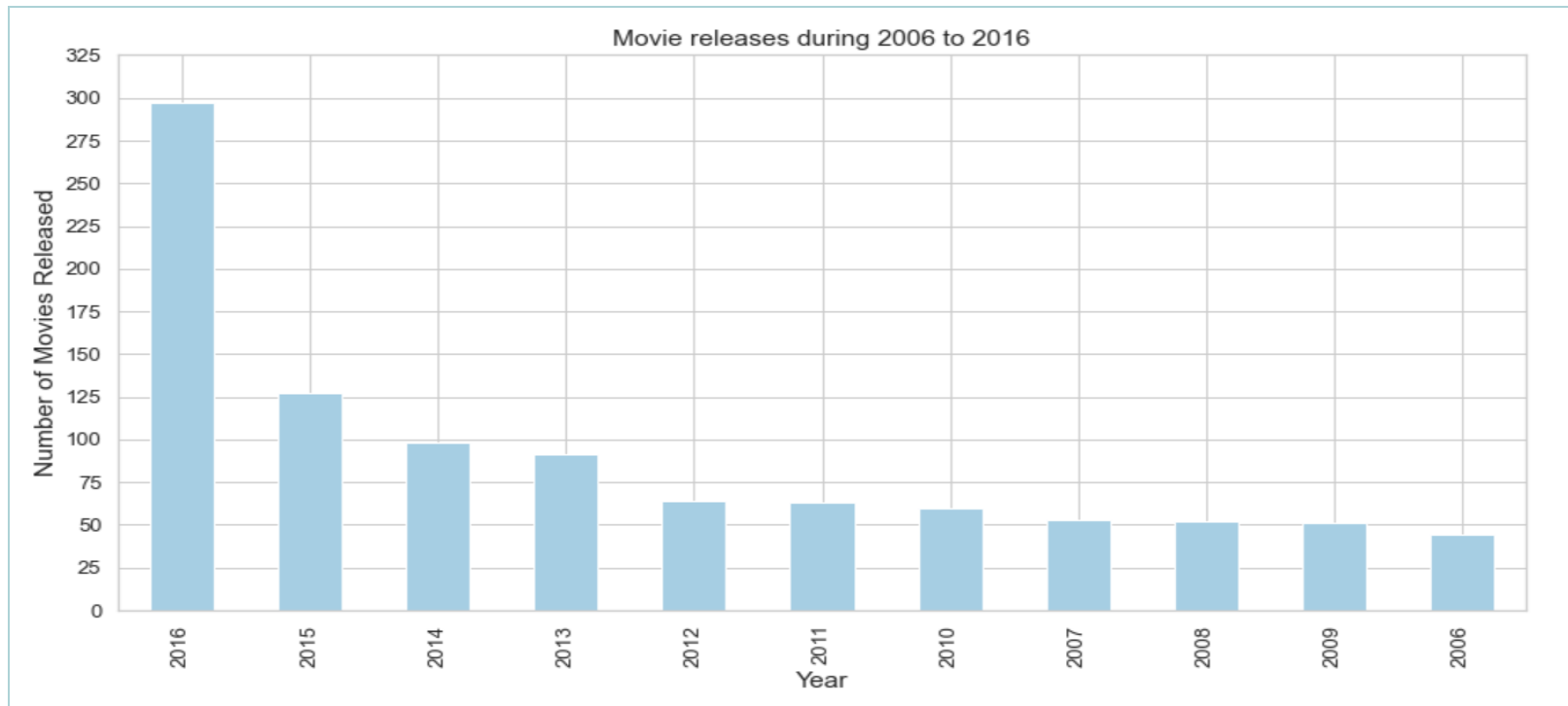
- Problem Statement
- Analysis Based on Yearly Metrics
- Analysis Based on Movie Revenue
- Analyzing the co-relation between all the columns using heat-map
- Comparison of movie 'ratings' with each parameter
- Analysis Based on various movie Genres
- Percentage of Movies & Genres
- Revenue & Genres
- Ratings & Genres | Votes & Genres
- Analysis Based on the Director of the movies
- Conclusion
- Actionable Insights
- Approach of the analysis

Problem Statement

- IMBD Ratings has been one stop for all the Movie Fanatics who like to do a tiny research by viewing the ratings to see if the movie is worth their buck to go to the cinema hall. However, this happens much after the movie has hit the theaters.
- The point to focus here is, would there be a way where we can predict movie revenue or its success way before a certain type of movie releases. This would significantly benefit the professionals in the Movie Business.
- Let's see if we can predict the movie success by doing an Exploratory Data Analysis (EDA) to the IMDB dataset of 1000 movies from 2006-2016 and by comparing various elements of movies.

Analysis Based on Yearly Metrics

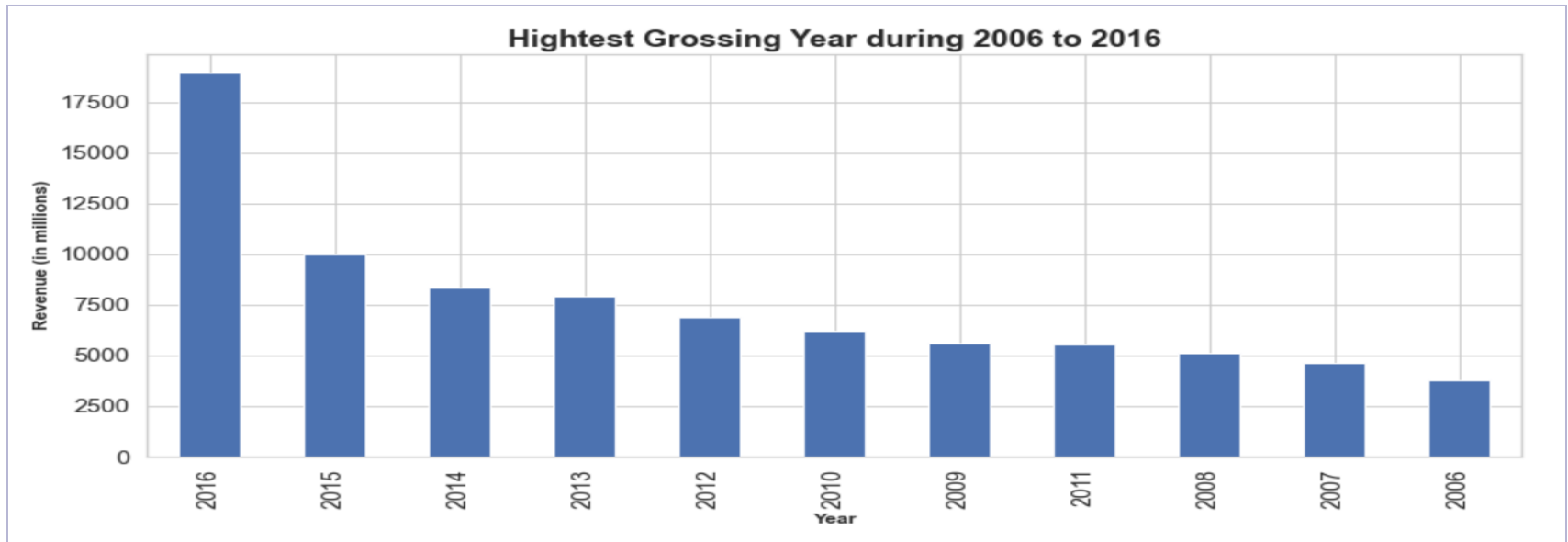
How many movies are released in each year? Which year saw the highest number of movie releases?



- Above chart clearly shows that **Year '2016' produced 297 movies**, which was the **highest number of movies** released during the **period of 10 years**.

Analysis Based on Yearly Metrics

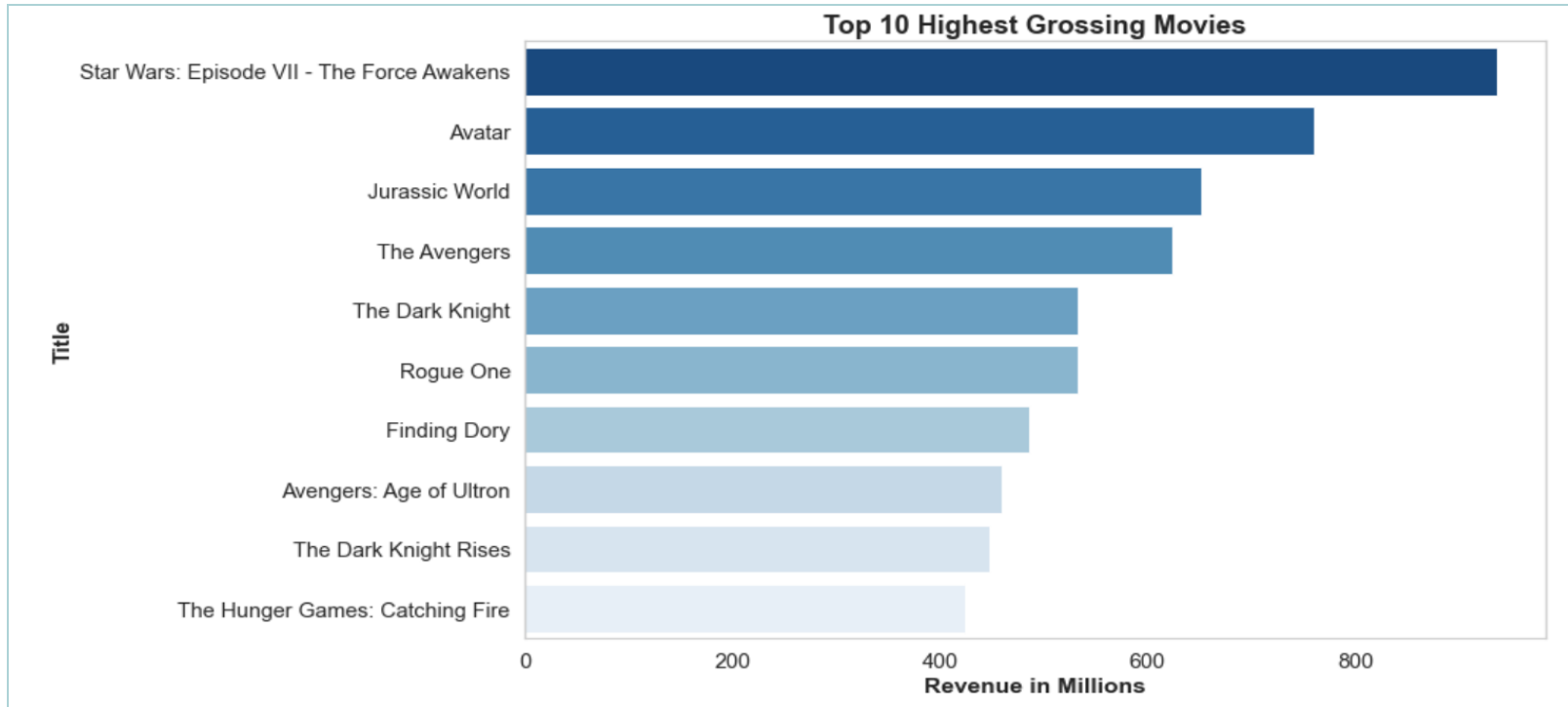
Which year generated the highest revenue?



- Similarly, the Year '2016' was the highest grosser of movies with USD 18926 MN (sum of revenue of all movies released in 2016) during the period of 10 years.

Analysis Based on Movie Revenue

Which are the Top 10 Highest Revenue making movies?



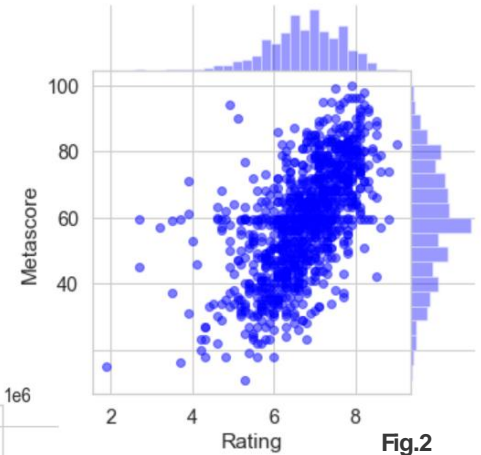
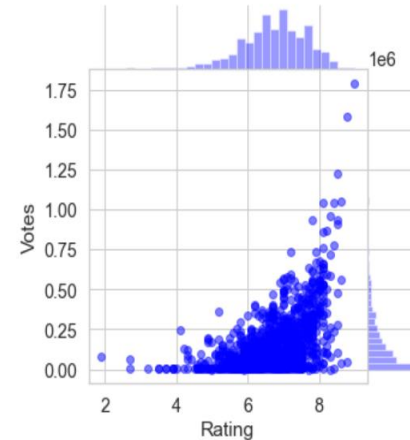
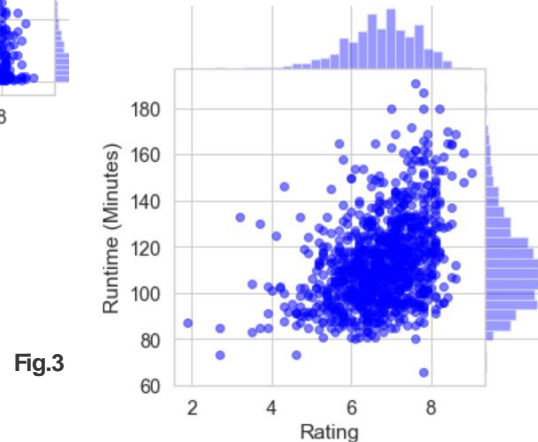
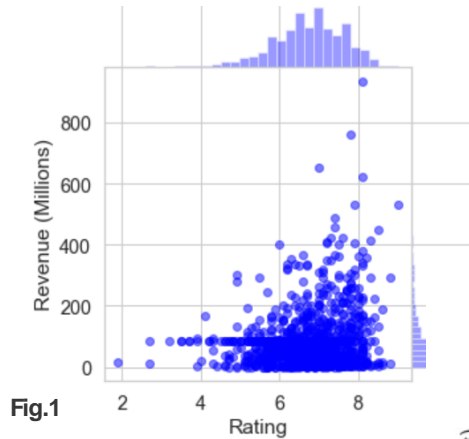
- Out of the Top 10 Revenue making movies; 'Star Wars: Episode VII - The Force Awakens' was the highest grosser of movies which made USD 936 MN.

Analyzing the co-relation between all the columns using heat-map



- The above Heat-map gives a complete overview of the co-relations with each columns.
- Movie 'Rating' shows positive co-relation with 'Metascore', 'Revenue (Millions)', 'Votes and 'Runtime (Minutes)'.

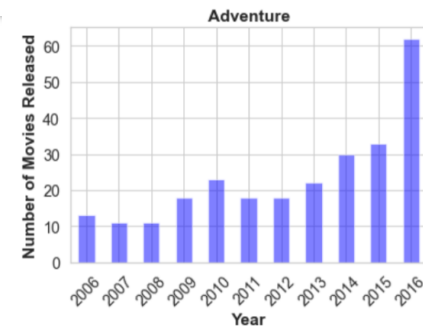
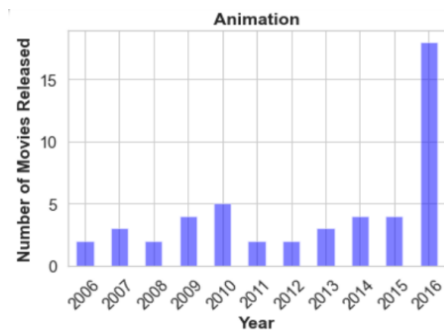
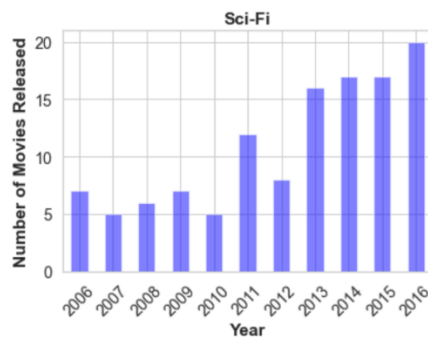
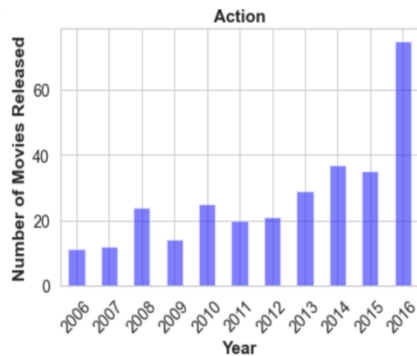
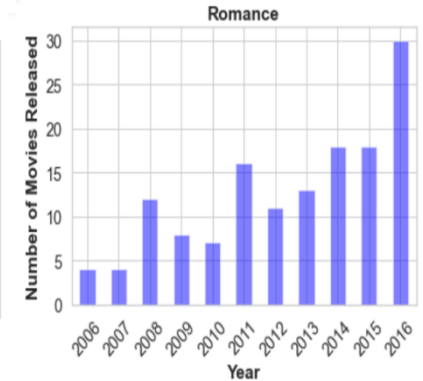
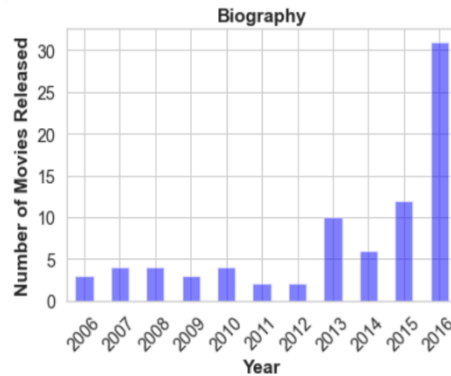
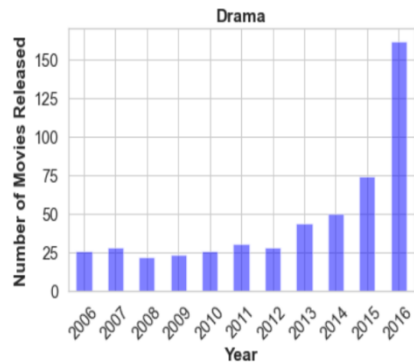
Comparison of movie 'ratings' with each parameter



- From the above charts, we can ascertain that **Metascore is highly co-related to Ratings. Audience rating was much close to the Metascore (Critic rating).** (refer: fig.2)
- We can also see that most **high grosser movies are depended on Public Ratings.** And, its clear that the movies that are **highly rated by the public have earned quite a decent amount of revenue share.** (refer: fig.1)
- However, there's only fair co-relation between the two. Hence, ratings cannot be the only source for a movie success but can certainly be one of the key aspects.
- It is also interesting to see that a **movie run-time** can also **significantly affect** for a movie to receive a **good rating.** (refer: fig.3)
- **Public votes and ratings are highly co-related to each other .** (refer: fig.4)
- Lastly, audience has given **large number of votes and ratings** to the movies that they have loved the most. Lower the run-time, higher the rating.

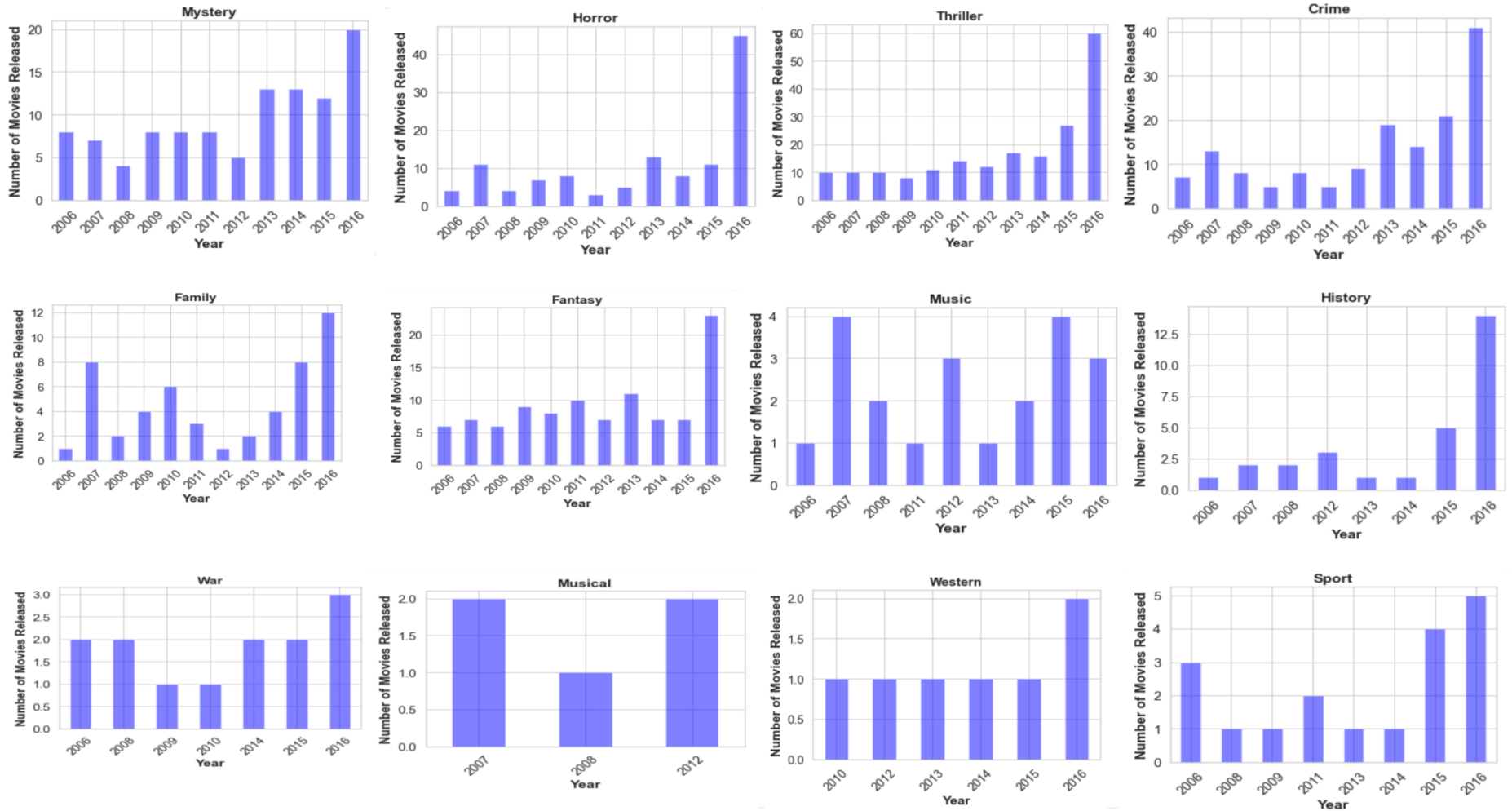
Analysis Based on various movie Genres

How many movies of each genre were released in each year?



- Movies Genres: Drama, Biography, Comedy, Romance, Action, Sci-Fi, Animation, etc are popular amongst viewers and was released every year through-out the span of 10 years. These Genres also saw highest number of movies made during this period.
- While few others were not produced regularly and were made rarely at the max 6-8 times in 10 years period.

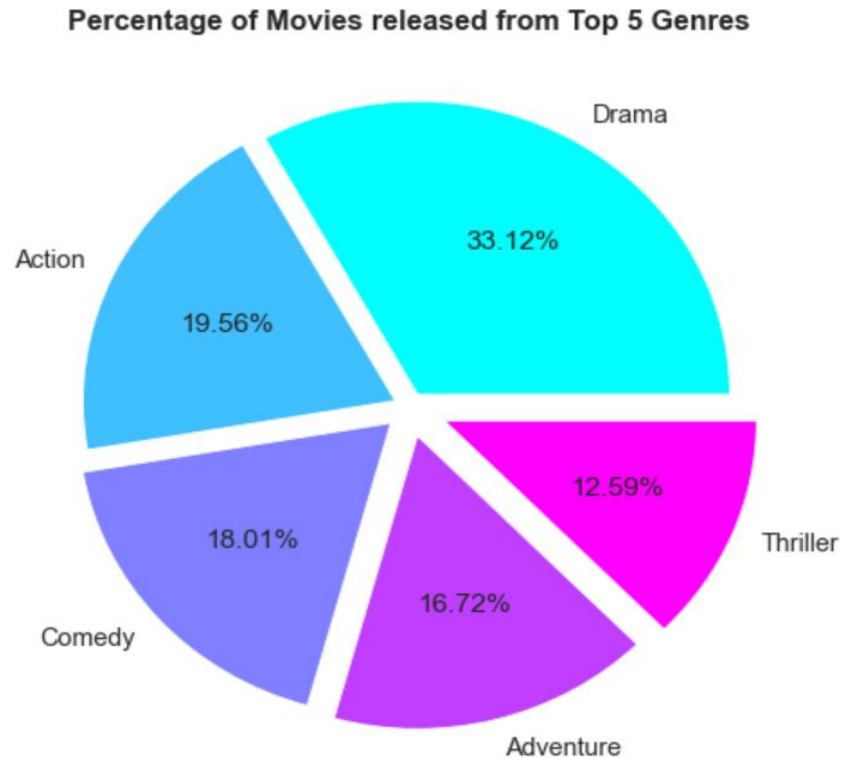
How many movies of each genre were released in each year?



- Movie Genres: Western, War and Sports produced least number of movies during this decade.

Percentage of Movies & Genres

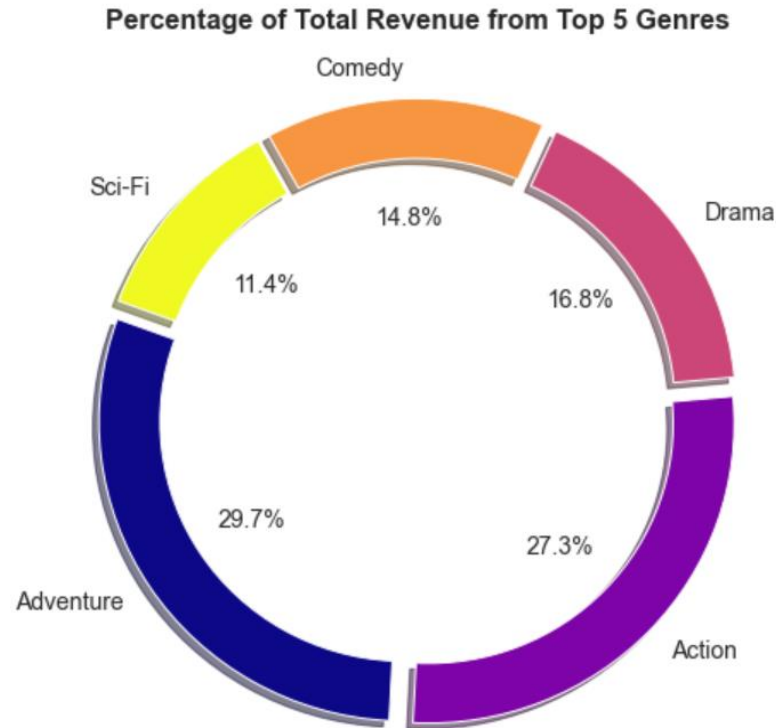
Determining the percentage of movies released in each Genre



- From the above pie chart, we can see that highest % of movies released are of Drama Genre which comprises of 33.12% of movies.

Revenue & Genres

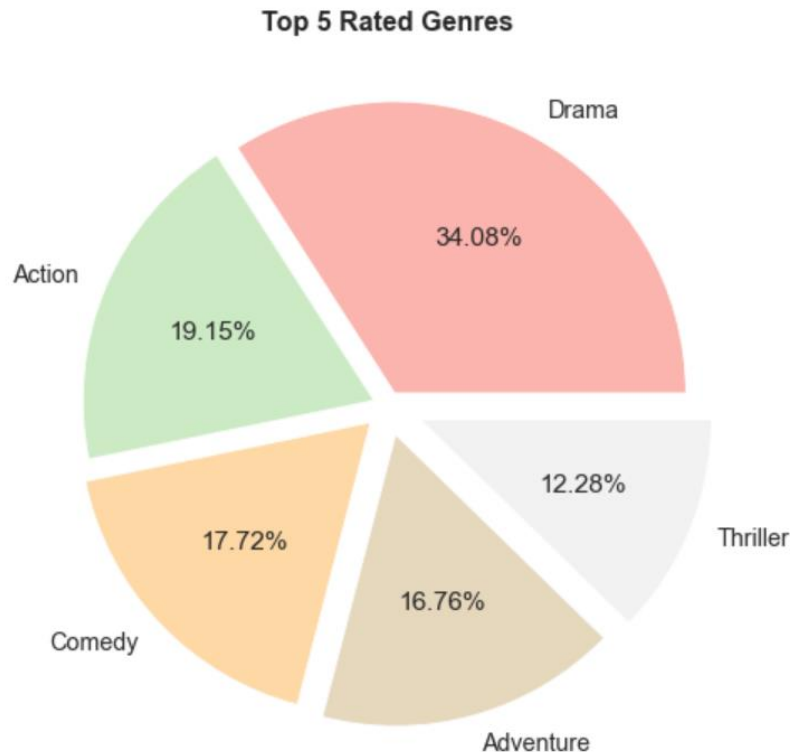
Determining revenue percentage received from each Genre



- The above pie chart clearly depicts that movies of '**Adventure**' Genre was the **highest grosser with 29.75%** of total revenue.
- '**Action**' Genre which was **27.26%** was much **closer to the highest percentage of revenue** making Adventure Genre.
- This concludes that **Adventure and Action Genre movies are highest grosser Genres** compared to the others.

Ratings & Genres

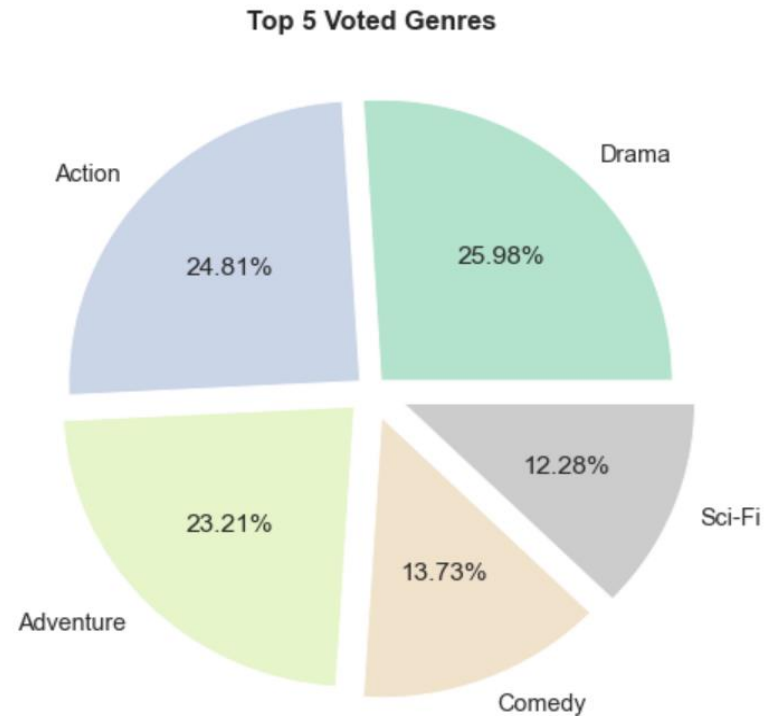
Determining IMDB Ratings received for each Genre



- The above chart shows us that the movies of '**Drama**' Genre was most popular amongst movie fanatics during this decade with **ratings of 34.08%**.

Votes & Genres

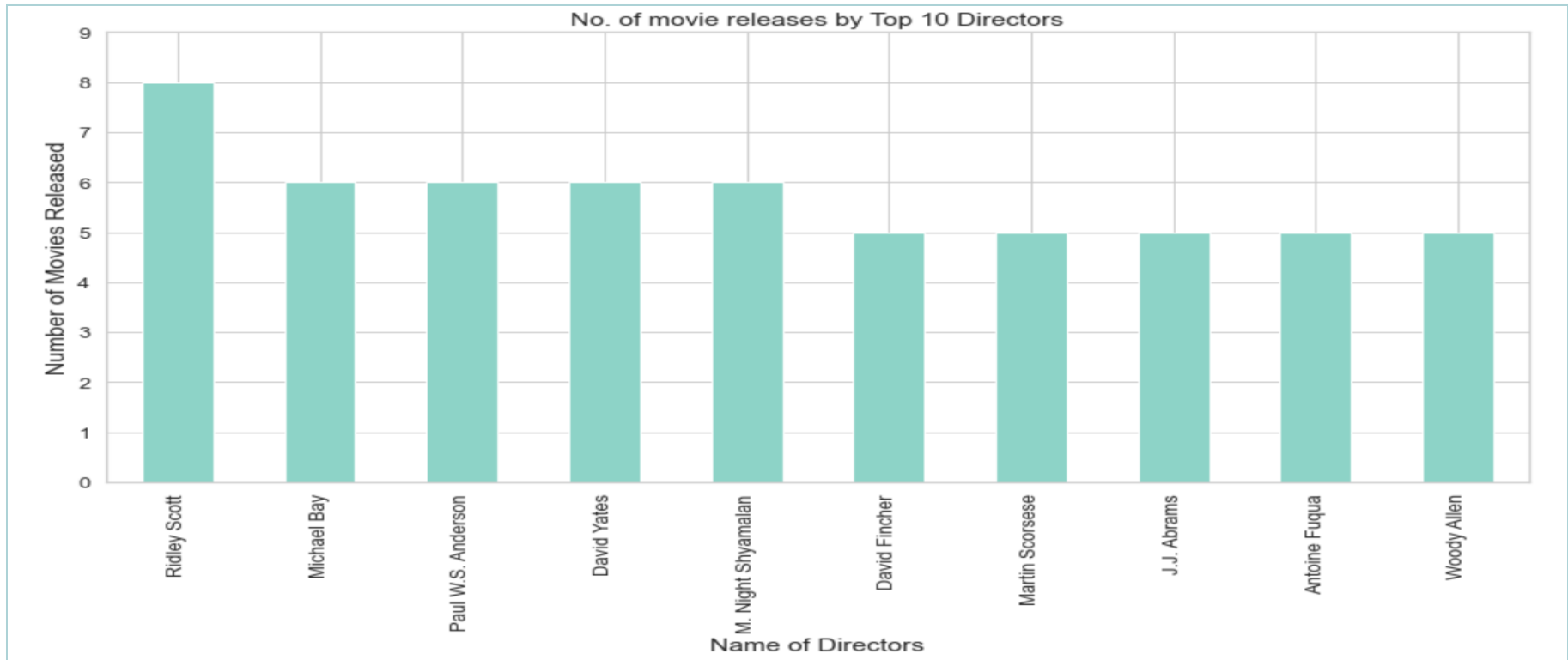
Determining Audience Votes received for each Genre



- Similarly as we saw with the Audience ratings, public **voted the highest for 'Drama'** Genre with **25.98% votes**.
- Action and Adventure** received **close enough votes as 'Drama'**.
- This shows that if the public appreciates a movie, they not only vote more, they also give a good rating.

Analysis Based on the Director of the movies

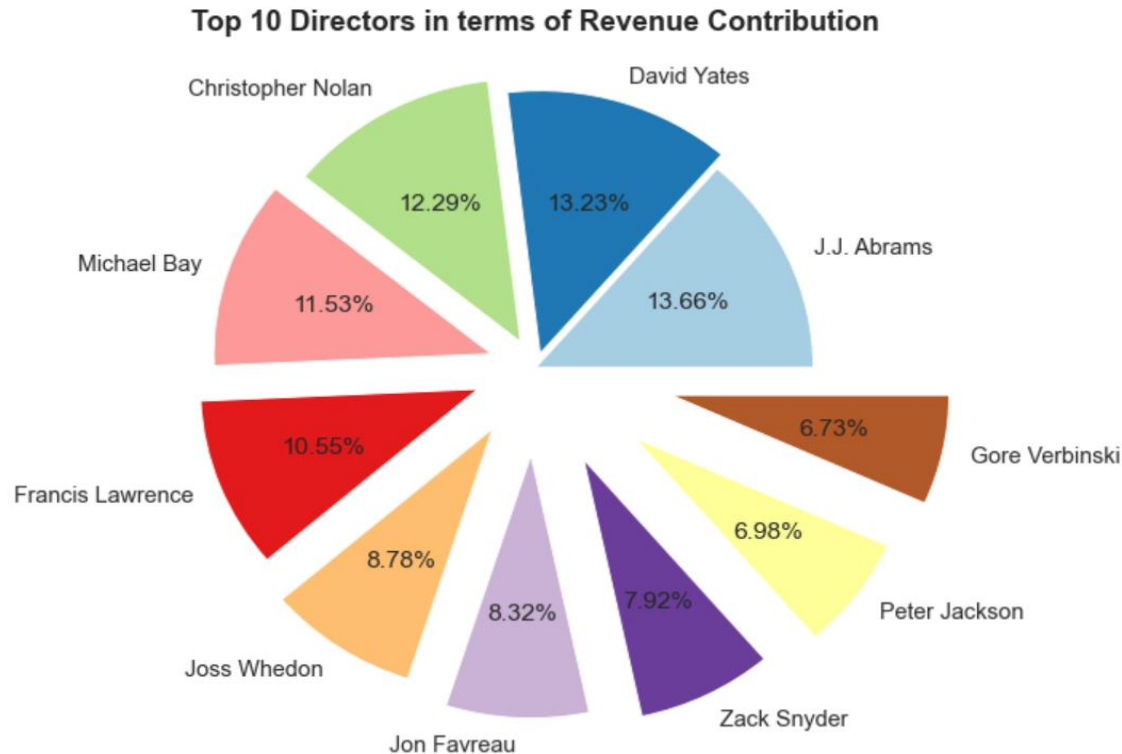
Which are the Top 10 movie Directors in terms of highest number of movies released?



- This bar chart clearly shows that out of the top 10 Directors, Ridley Scott has directed highest number of movies, .i.e.8 movies during this decade.

Analysis Based on the Director of the movies

Which are the Top 10 movie Directors in terms of Revenue contribution?



- From the above pie chart, we can infer that, **JJ Abrams** with **13.66% Revenue share** has Directed the **highest grosser movies of the decade**. One of them being, '**Star Wars: Episode VII - The Force Awakens**' which we already saw was the **highest grossing movie of the decade**.
- However, **David Yates** was much **closer to JJ Abrams** with **13.23% revenue share**.
- It is also observed that, though Ridley Scott has topped the charts for making maximum number of movies during the decade, however, his revenue share is not amongst the top 10.

Conclusion

Summarizing the above analysis and their findings:

- Year 2016 produced highest number of movies and earned highest revenues during the decade of 2006-2016.
- Out of the Top 10 revenue making movies, 'Star Wars: Episode VII - The Force Awakens' topped the box office.
- Audience ratings of movies (Ratings) were quite close to the Critic ratings (Metascore).
- The more the public appreciated a film, the more they voted and gave a good rating.
- Lower the run-time, higher the rating.
- Movies Genres; Drama, Biography, Comedy, Romance, Action, etc was released every year in this decade.
- Highest number of movies were made 'Drama' Genre.
- In terms of revenue, Adventure Genre made more money than other genres, however, Action Genre made close enough.
- Audience rated Drama genre the highest of all.
- Though the Directors who made higher number of movies, they did not earn high revenue share.

Actionable Insights

- After conducting a detailed analysis on the data with various graphs and comparisons various elements, this is to conclude that, IMBD does have various other factors (like Movie Genre, Metascore & Movie Run-time) other than 'Ratings' which has significant influence for a movie to become a chart buster/high grosser.
- It would prove highly beneficial for movie business professionals to utilize these parameters such as **Genre, Metascore, Public Votes and Movie Runtime** to predict the box-office success of movies.
- Further to predict the success factor of a movie, a '**Recommendation Engine**' can be built to capture the audience's past behavior towards similar type of movies/movie genres and other elements.

Approach

- **Define Problem** – After studying the data-set and researching fairly on movies in IMDB, arrived at the problem statement.
- **Choosing Right Tools** – Imported all the required packages such as NumPy, Pandas, Itertools, Matplotlib and Seaborn for visualization purposes.
- **Data Collection** – Imported 1000 movies dataset from GitHub.
- **Pre-profile** – performed pre-profiling on the dataset.
- **Data Pre processing** –
 - Removed Description column,
 - Replaced null ('NaN') values in Metascore & Revenue with median & mean respectively.
 - Checked for duplicates
- **Post-profile** – performed post profiling on the modified data set.
- **Asking right Questions** –
 - Listed down questions post-profiling observations,
 - Analyzed them with various pandas functions, NumPy arrays,
 - Percentage calculations and plotted graphs to infer observations.
 - Standardized columns for better analysis and comparison purpose.
 - Comparisons with various movie elements derived findings.
- **Conclusion /Summarization** –
 - Summarized the above EDA analysis with findings.
- **Actionable Insights** – Derived actionable insights in accordance to the findings on the detailed analysis



Thank You!