

# **BIG DATA MANAGEMENT AND ANALYTICS**

ARAVINDAN SRINIVASAN

2981707

Aravindan.srinivasan@student.griffith.ie

## **APPLIED DATA ANALYTICS**

---

### **EXPLORATORY ANALYTICS AND VISUALISATION ON ART OF MOTION PHOTOGRAPHY**

---

A movie is created with motion picture camera by taking photographing scenes. The idea of cinema is to communicate various ideas, stories etc. One of the popular entertainments is cinema. It is a powerful tool for communication of important ideas around the world. Movies are getting enjoyed around the world by dubbing or subtitles to other languages. Movie makers are profited by highest number of audiences watching the movies. Developing a good movie is not enough, **predicting the future and understanding the needs of audiences** and giving a right movie is important. Let's do some analysis to fetch the important answers for crew members in movie industry, critic reviewers and audiences. The dataset we are using for analysis is a CSV file("movie\_metadata.csv").

#### **About the dataset:**

**Dataset link:** <https://www.kaggle.com/carolzhangdc/imdb-5000-movie-dataset>

The datasheet is downloaded from Kaggle website. This dataset consists information of 5043 movies, 28 attributes and movies around 66 countries. It contains 2000+ directors and 1500+ actor and actresses. In this dataset, our class variable is "imdb\_score" which contains the scores given by imdb for movies around the world. The scoring pattern of imdb works by fetching all the individual registered votes and use them to score movies. They use algorithm to mark the movies. Imdb have never given their method for scoring the movies, but they use the same method for all the movie scores.

Listing out the attributes and their description:

Attribute	Description
movie_title	Title of the movie
color	"Black and White" or "Color"
language	Languages like English, Hindi etc
country	Movie released in which country
budget	Movie budget in Dollars (\$)
gross	Earnings made by movies (\$)
duration	Duration of the movies in minutes

<b>plot_keywords</b>	Keywords for movies description
<b>genres</b>	Film categorization like comedy, drama etc.
<b>movie_imdb_link</b>	Imdb link of the movie
<b>title_year</b>	Movie released year
<b>facenumber_in_poster</b>	Poster consist of number of actors
<b>aspect_ratio</b>	Aspect ratio of the movie made
<b>imdb_score</b>	Score given by imdb website
<b>content_rating</b>	Depicts which group is suitable to watch movie
<b>num_critics_for_reviews</b>	Number of reviews given on imdb
<b>num_voted_users</b>	Number of votes for the movie
<b>num_user_for_reviews</b>	Number of reviews about the movie
<b>director_name</b>	Name of the director of the movie
<b>actor_1_name</b>	Lead role in the movie
<b>actor_2_name</b>	Second lead in the movie
<b>actor_3_name</b>	Support actor in the movie
<b>director_facebook_likes</b>	Likes got by the director in Facebook
<b>actor_1_facebook_likes</b>	Likes got by the lead role in Facebook
<b>actor_2_facebook_likes</b>	Likes got by the second lead role in Facebook
<b>actor_3_facebook_likes</b>	Likes got by the support role in Facebook
<b>cast_total_facebook_likes</b>	Total number of likes by the crew
<b>movie_facebook_likes</b>	Number of likes got by the movie in Facebook

### **Problem Statement:**

There are various factors which make a movie more successful compared to other movies. Therefore, we are going to do some exploratory analytics on this dataset to get the insight of the data. We will be showing the visualization on the dashboards using Tableau. Focussing on predictions by analysing various attributes in the given dataset. These results will help film companies, actors, critic reviewers and audiences to spend their money efficiently on movies.

### **Exploring the Data using Tableau (Dashboard):**

#### **1<sup>st</sup> Dashboard**

##### **Target Audience:**

**Morgan Freeman** is an American actor. We will be doing some analytics on his past information and do some prediction on choosing him the best genres for his next movie.

##### **Question:**

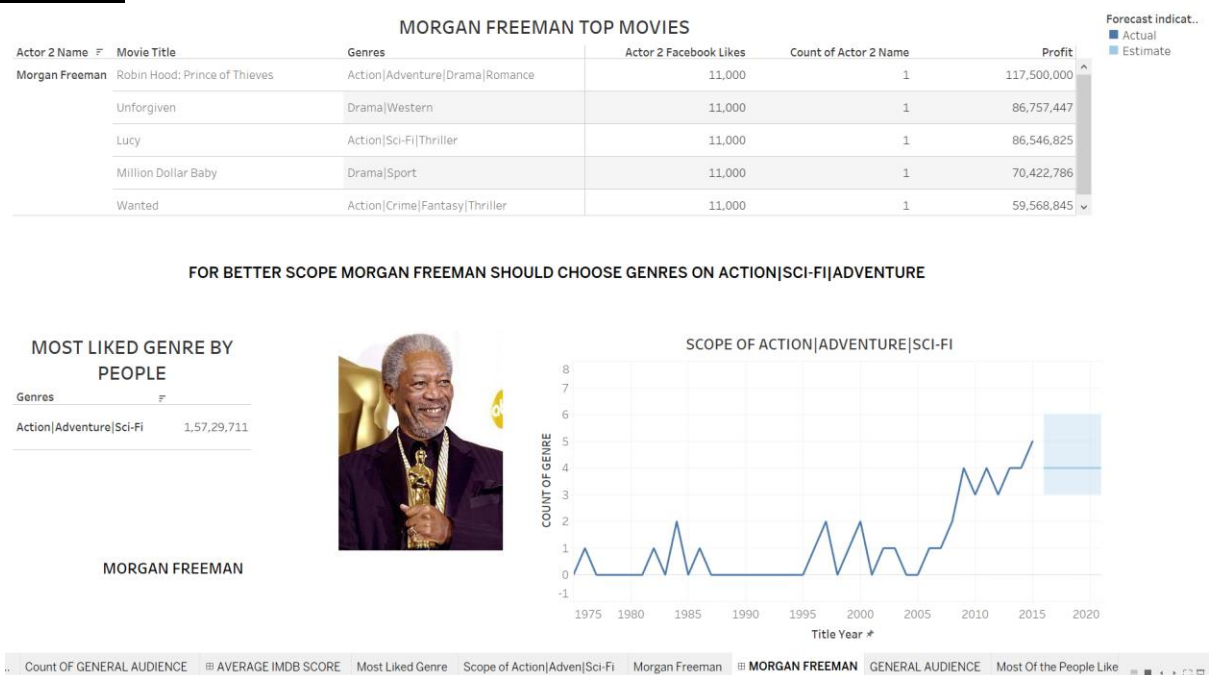
Which genre movie does Morgan freeman have to choose for his next movie?

##### **INSIGHTS:**

Initially, at the top of the slide we presented Morgan's top 5 movies. His major success on the Action genre movies. He made a huge profit on the movie "Robin-Hood" which involves

action, adventure, romance and drama. In the **image**, we fetched the most voted genre from various audiences around the world through Facebook. It is clear that, people worldwide want to see Action, Romance and Sci-fi. Most important part is predicting the scope of Action, Romance and Sci-Fi genre. Many movies are going to be released on this particular genre. His number of likes in Facebook is 11,000. His advantage is exploring the various genres. But this cannot support every time. In order to improve his Facebook likes, he has to choose the genre that is liked by most of the people. Therefore, he has to select the genres of action, adventure and Sci-Fi journal to have an increase likes in Facebook as well as confirm hit over the next movie. Moreover, the scope of this genre is increasing every year. This is the right time to commit this journal and might break his own record in Facebook as well as his salary in the next movie. It is the right time to create his spot for his next Oscar award on this journal.

## DASHBOARD:



## 2<sup>nd</sup> Dashboard:

### Target Audience:

**Director and Producer** are the captain of the ship. Without them, a movie cannot have a proper structure. This dashboard is specially for them to focus on which genre they have to choose for their next movie.

### Question:

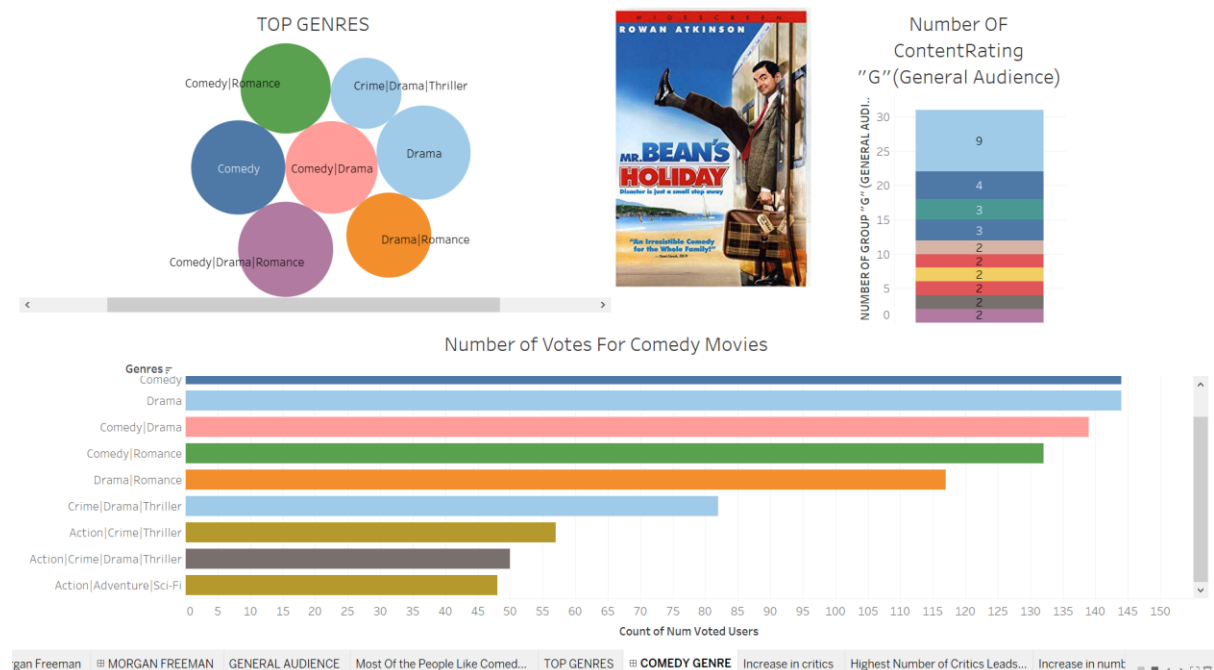
Which genre is liked by most of the people?

### INSIGHTS:

In the image2, we are looking at top most of the genres used by various directors. Most of the common average genre is comedy. Every people enjoy the comedy movie. Next to that drama and thriller is the expected genre from people. In this, we are looking the most voted genres

from various people. All the movie audiences cast their vote on comedy genre (147 votes). Next to that drama has the maximum number of votes (144 votes). So, people around the world expect comedy and drama together. This will create a huge impact among the audiences. The most important reason for a successful movie is having highest number of audiences watched the movie. Group “G” is the general audiences where all the people are allowed to watch the movie. If there is different content rating, particular age group are not allowed to watch the movie. For this reason, comedy genre has the advantage of having all the audiences to enjoy the movie. For instance, Mr bean movie budget was very low, but the gross was high compared to the budget. This movie does not have a great imdb score, but the gross was high. It is because of two reasons.

- i. It has the “G” content rating, which means general audience. All the age group are permitted to watch the movies. Whole family got the chance to enjoy a movie.
- ii. Another reason is most of the people like the comedy journal. That is one of the main reasons why it has a high grossing. 144 votes casted for comedy journal which is the highest compared to the other journals.



### 3<sup>rd</sup> Dashboard:

#### Target Audience:

**Film criticism** is analysing and evaluating movies around the world. This dashboard will help the people who do film criticism as a profession. And it has advantage of making a movie popular worldwide. It will be helpful for the **whole crew of the movie**.

#### Question:

How to attract audiences and make a movie popular?

#### INSIGHTS:

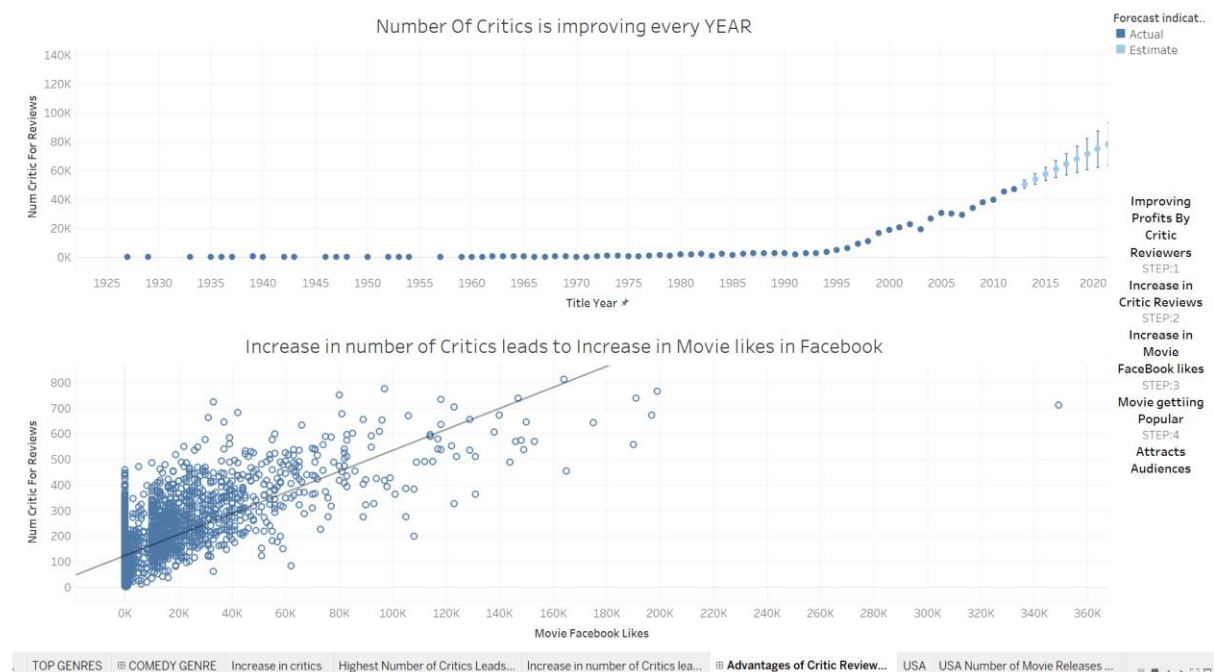
First slide depicts the increase in number of critics every year. **“FUTURE FORECAST”** estimates number of critics will double up in future. Reason for this is, people started to make

their own criticism on movies and posting that reviews on internet. If that particular post has more views, then those people who made criticism are getting paid by various web sources. For instance, video posted on YouTube and getting paid by YouTube organization. Second slide explains the movie likes. The more critics a movie get, leads to more likes to that movie on Facebook. This will help the film crew by their movie getting popular.

Let us take a sample movie Avatar and steps followed here is:

1. Increase in number of critics of a movie Avatar,
2. Leads to increase in movie likes on Facebook
3. The more the movie get likes, getting an eye worldwide.
4. Audiences getting attracted and watch the movie Avatar.

These are steps surely going to happen if it has lots of critics. To gain a greater number of reviewers, a preview show can be provided to the critic's reviewers. But it has a little bit of risk. If the movie is not that up to the mark, it will lead to a big flop. If the producer and director are sure about the movie strength, then they can go ahead. Because it might create a hype to the audiences and reaches greater number of audiences.



## 4<sup>th</sup> Dashboard:

### Target Audience:

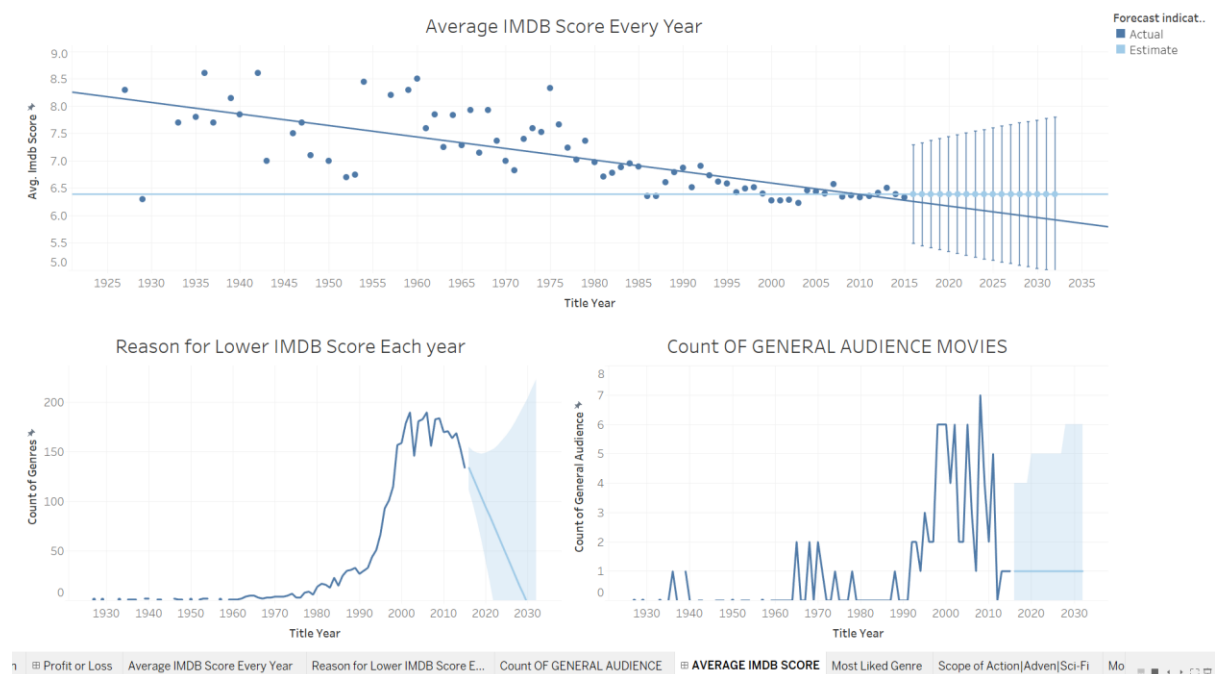
**Director and Producers** are the ultimate user for this dashboard. Every year the average imdb score is reducing. The lower the movie score the lesser the viewers will be for the movie. For a successful film, imdb score is one of the important factors. This dashboard will list out the reasons for reduced average imdb score each year.

### Question:

How to avoid getting a lower imdb score?

### INSIGHTS:

Trying out a new genre might make a movie in risk in success. If a director and producer want to give a successful movie, they have to follow the path in past. Increase in genres each year leads to reduced average imdb score each year. This is because, if a director wants to cover up more genres on a single movie. He fails to cover all the depth of the genre. Trying out a new genre, have the chance of not getting content rating of “G” (General Audiences). Automatically, audiences will be filtered based on age. Creates a problem on the profit of the movie, which effects the whole crew. If they choose to make a movie using many genres on a single movie, it needs a lot of work. For instance, meet the Spartans is a spoof movie of 300. This movie covers various genres and lead to huge profit to the production. This created an impact on covering on many genres. Followed by that, many movies have faced the flop following the footpath of meeting the Spartans. On the basis of average, it is advised to take a few genres and cover the depth of that particular genre. So, the production unit and the director can have a peaceful mind during the release time of the movie. If they follow on using multiple genres on a single movie. The prediction will be true. And it makes the audiences sick towards the movies.



## 5<sup>th</sup> Dashboard:

### Target Audience:

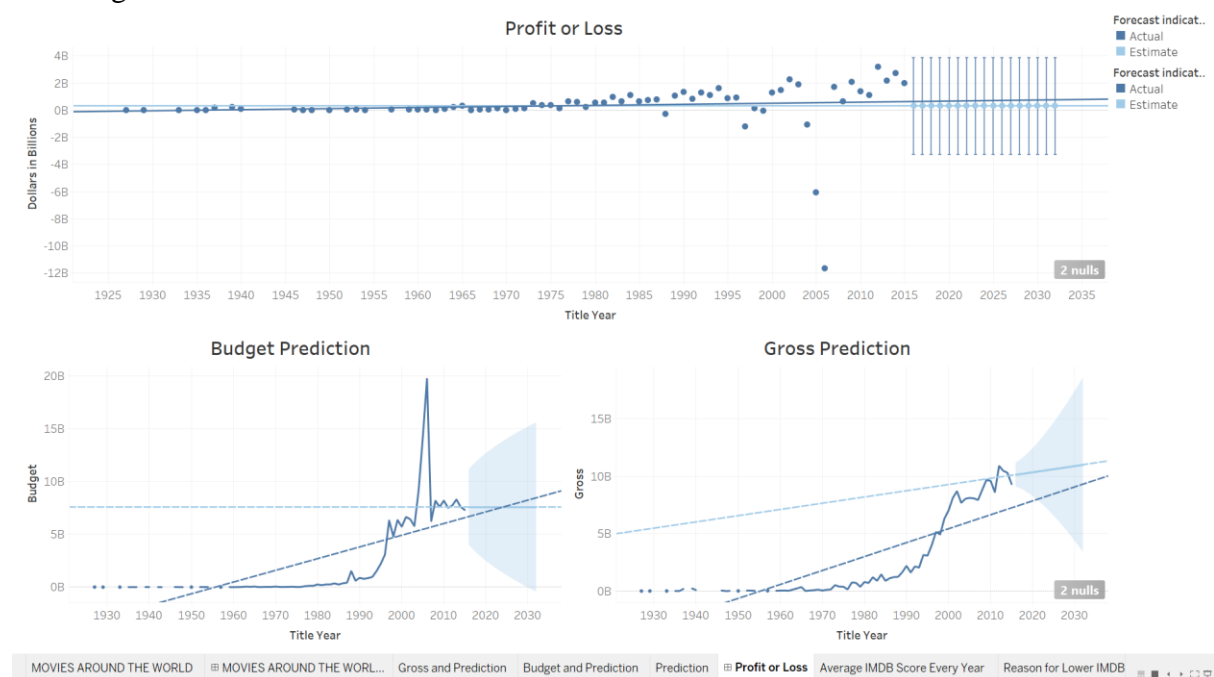
**Producers** are the key for all movies. They listen to the script from the directors and accepts to produce the movie. But they have no idea, how much money they have to invest on movie (Budget). This dashboard will help them to decide how much budget they have to invest and even predict the upcoming budgets in future.

### Question:

How much budget should I invest in my next movie and approximately how much gross can I expect?

### INSIGHTS:

Budget prediction slide will help the producers on how much money he has to invest on 2020. The following graph displays all the average budget of each and every year. This graph displays the information of past as well future (1930-2032). For the year 2020, the average budget will be 7.5Billion\$. Gross prediction graph will help the producers on how much gross he will get in the following years. Same like budget prediction, gross prediction will display the average gross of every years. This graph depicts the information of past as well as future (1930-2032). For the year 2020, the average gross will be 10.3Billion\$. Final graph will display all the information of the movie average profit made on every year. It displays the profit or loss acquired on every year. Having the information on entire movies leads to predict the future of every year. This budget prediction is very important for the production unit to have some idea on producing a movie. Moreover, gross prediction is also very important. This is where the **other dashboards are linked**. This dataset will be linked to many dataset on the relationship by having the budget in this dataset and choosing the genre on different dashboard, viewing the steps on making the movie popular. Thus it has a greater advantage for the whole film crew.



## 6<sup>th</sup> Dashboard:

### Target Audience:

Number of movies released in USA is higher compared to other countries. This is because, English being a spoken all around the world. USA country have all the better technology which attracts new directors to try different things in movies. **Directors** are the user for this dashboard.

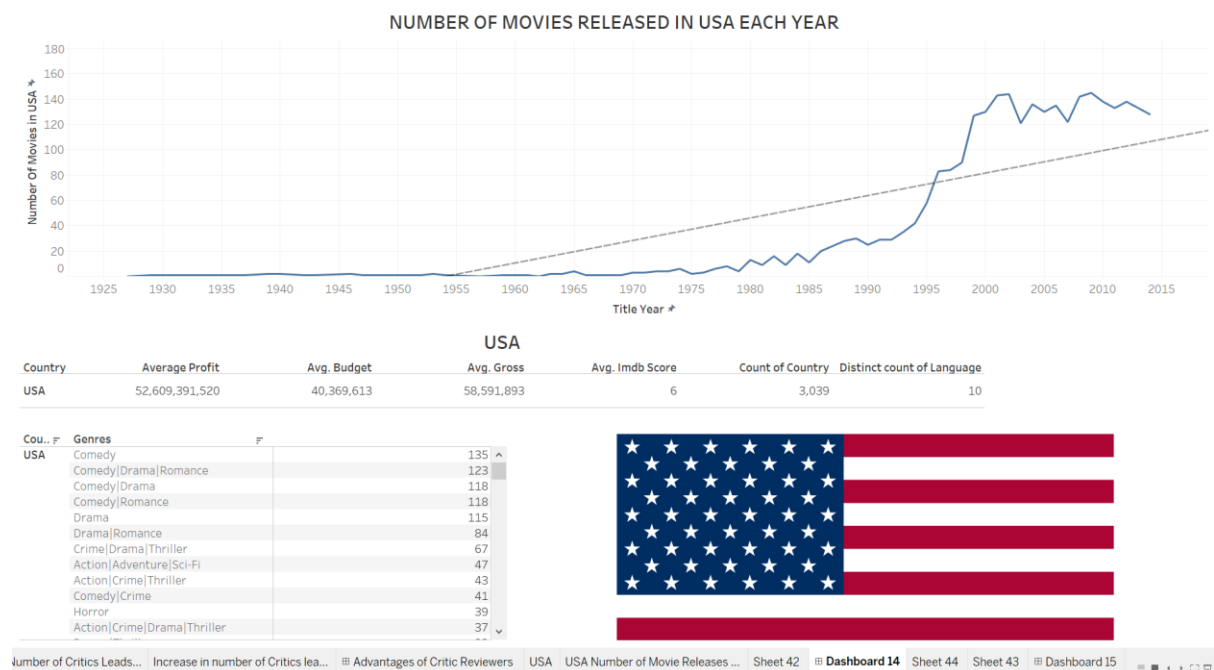
### Question:

In which language and country, I should direct my upcoming movie to have a worldwide audience?

### INSIGHTS:



The first slide depicts number of movies released every year. This will encourage a new director to commit a new film in USA country. Second slide displays all the basic movie information in that country. It lists the average of profit, budget and profit. Approximately, 3000+ movies released till now. Third slide displays, the most used genres in that country. Most of the USA films are comedy and action-based movies. This gives a prediction for the directors to choose a genre and view all the past information on that particular genre. Most of the movies are made with technology. Some action scenes are made with 3D effects which is enjoyed by various people worldwide. Moreover, USA is the only country who have lots of technology which will be useful and easier for the production unit and director to give a good movie for the audiences. All the movies made in this country are English, Since it is enjoyed around the world by using the subtitles or dubbing in their respective country and respective language. Having the common information of the USA country has the great impact on various directors, production unit, actor and actresses. Even the trend line depicts that more movies will be released in USA has the higher rates compared to other countries. If a movie of comedy genre with technology used, it has a greater impact on movies.



## **7<sup>th</sup> Dashboard:**

### **Target Audience:**

This dashboard will be helpful for the directors and producers. Listing all the genres by the profits. So, before release he can predict how much gross it will make.

### **Question:**

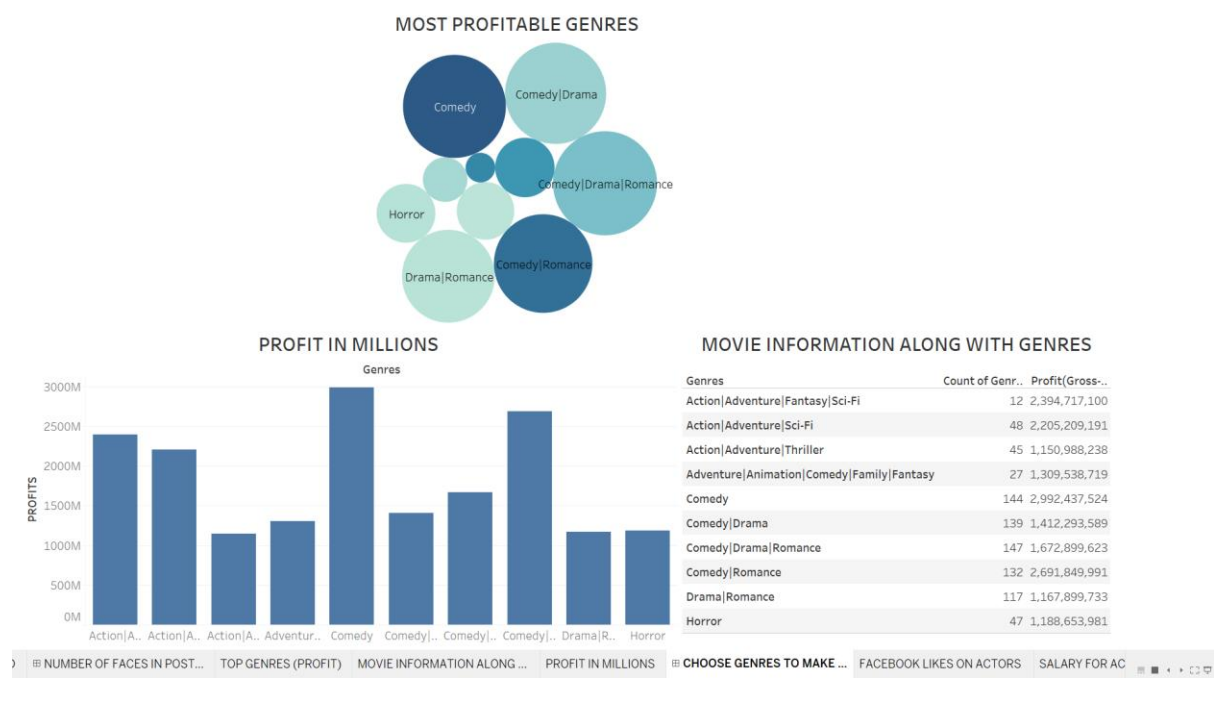
Which genre should I use to make more profit?

### **INSIGHTS:**

Initially, we listed top genres and the average profit made by the movies. For a director and producer, making a good movie is not enough, making money out of it is the main concern.



So, they have to analysis on which genre they have to choose to make huge profits. Audiences will enjoy the comedy movie, but a Sci-Fic movie will make them to watch the movie two to three times to understand the core of the story in that movie. This motivates directors, to create a complex story in an interesting way for audiences to make the movie profitable. For instance, **INCEPTION** movie, is one kind of a movie which has the genres of action, sci-fic etc. To be successful, experience is the major key for achieving greater heights. By following the information given in the dataset will help the whole crew on selecting the genres. Because we have listed all the successful movies on based on genres. We calculated the profit by subtracting the budget from the gross. So, this gives a idea about making a movie and expect how much gross it will lead to the whole movie making team. Therefore, comedy is the genre where it is liked by many people, moreover the average profit made by the comedy journal is higher compared to the other journals. You can view that 48 times the j=genre action, adventure and Sci-Fi is used. Some of the movies are avatar, avengers etc. These movies understand the liking of the people and given the right movie at the right time.



## 8<sup>th</sup> Dashboard:

### Target Audience:

Critic reviews will make use of this dashboard. This dashboard helps them to give an idea of good movies releasing around the world. It also gives the chance for them to review all the other languages movies around the world.

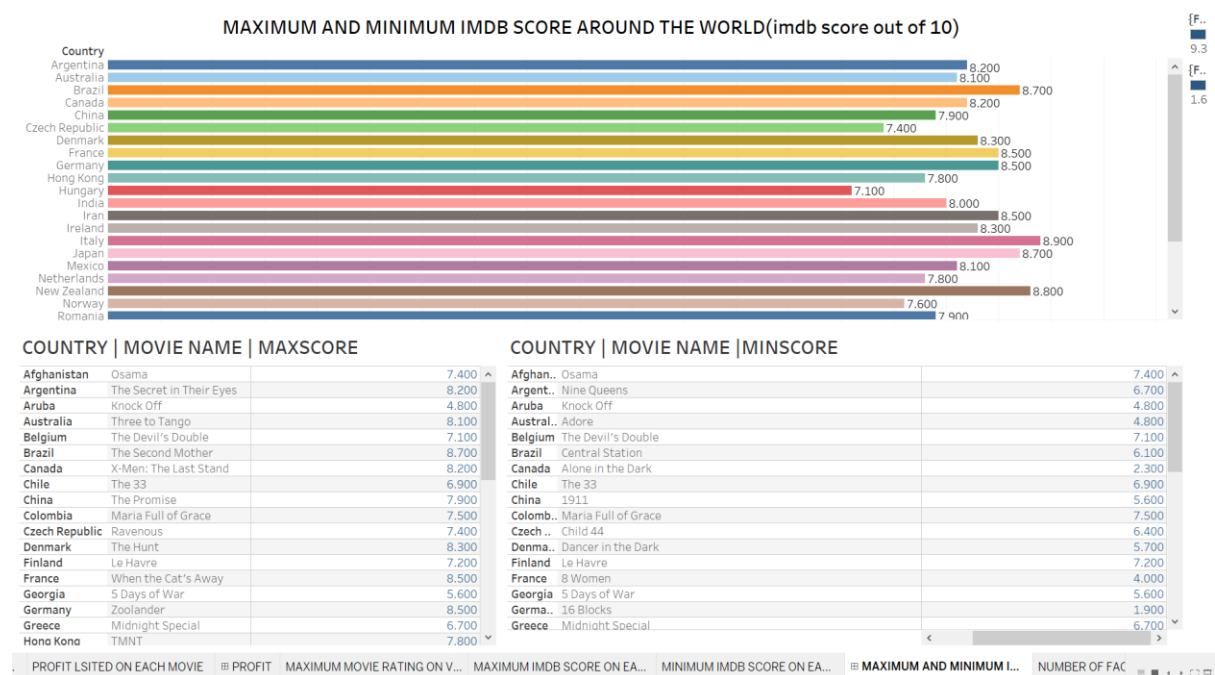
### Question:

List me all the countries highest and lowest imdb score:

### INSIGHTS:

In our first slide, we displayed all the countries average imdb score. So, it helps them to track which country has the good average imdb score. It helps other actors around the world to watch and learn acting skills on the good movies. In the second slide, we are displaying the

highest imdb score of every country. In the third slide, we are displaying the lowest imdb score of every country. This information is very useful, just like same as the previous dashboard. There are many dashboards which have the link to the other dashboards. This dashboard contains all the highest imdb score and lowest imdb score around worldwide. Therefore, if a director is going to choose a movie, he has all the information to **choose genre, how much budget to invest, how many face numbers he has to put in his poster and selecting the runtime of his movie etc.** This gives a strong relationship with the other dashboards. Even this one dashboard has the information of the exact value of fetching the average score of each and every country. To fetch the highest and make the decent movie on the safer side. He has to choose the country carefully in order to obtain the good imdb score. Even the registered imdb members will cast their vote and score of movies releasing on their country. It depends on the people. If all the people in an particular country hates the movie, which leads to the lower score in imdb.



## 9th Dashboard:

### Target Audience:

**Audiences** who wants to track the black and white movies in every year will make use of this dashboard. This dashboard displays all the information on movies whether it is black and white or color movie.

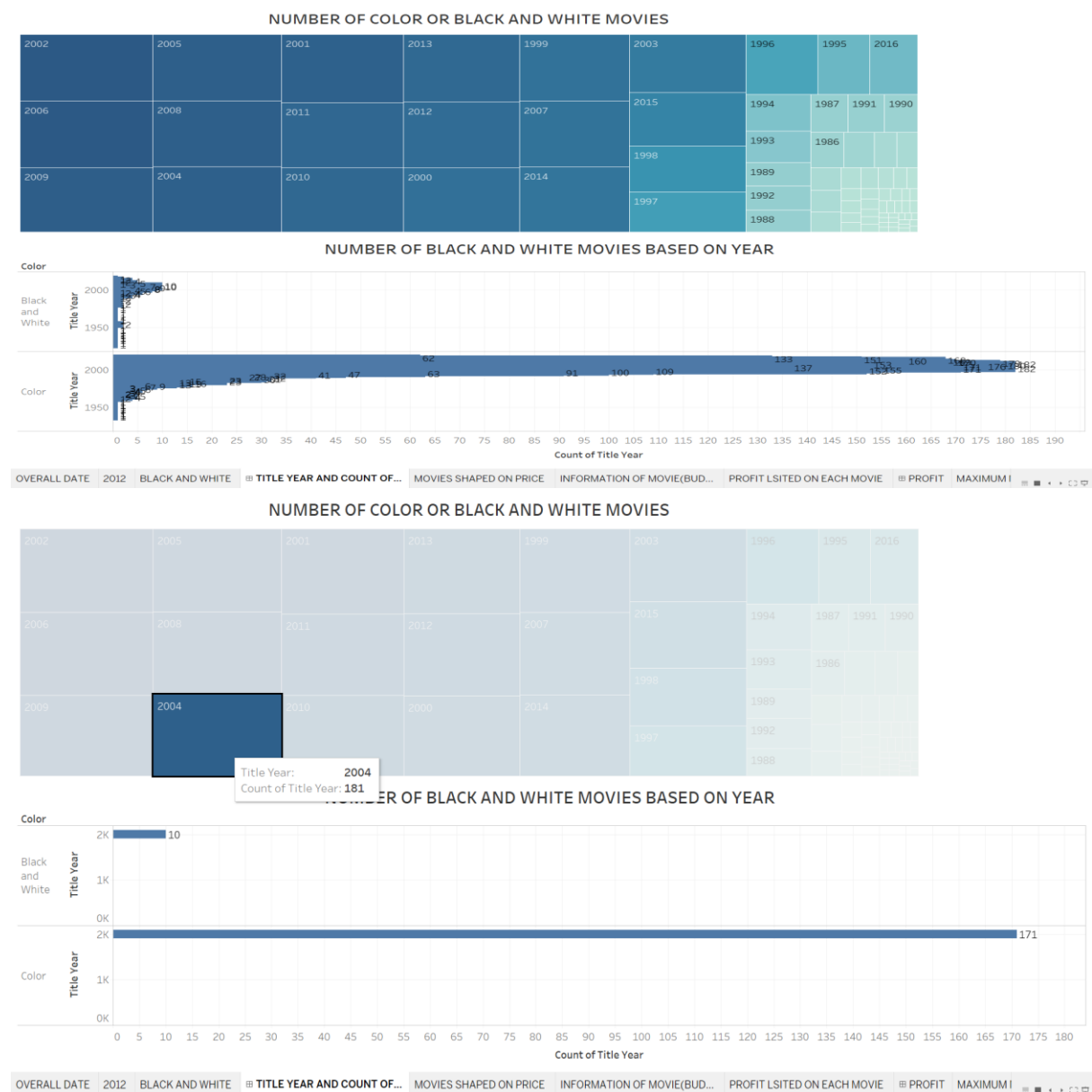
### Question:

List me all the black and white movies on every year?

### INSIGHTS:

In our first slide we displayed all the movie released in each year and grouped it based on the year. So that audiences can keep track of all the movies on each and every year. In the next slide if a user selects a particular year, our valid information is displayed by filtering the

noise information in the dashboard. Audiences who enjoy old movies, can track all black and white movies released in each year. Even now few black and white movies are released. Since we have the most liked genre as comedy. It gives a chance to choose a best actor and direct a movie like Charlie Chaplin. This will be dependent on director and actor. Because the budget is not important for this movie. Only pure talent skills are needed for making this movie a great success. This will create a trend on black and white. So, a director has to study all the work done by Charlie Chaplin and do some homework and write a script in his own way to make this movie a great hit. Apart from this, as a supporting factor, present world has greater technology to make the movie more and more beautiful. In this way a chance to hit the record of most grossing in black and white movies and create a trend set for the younger and new directors who are coming to the movie industry. This will also help the production team a great profitable business.



## 10<sup>th</sup> Dashboard:

### Target Audience:

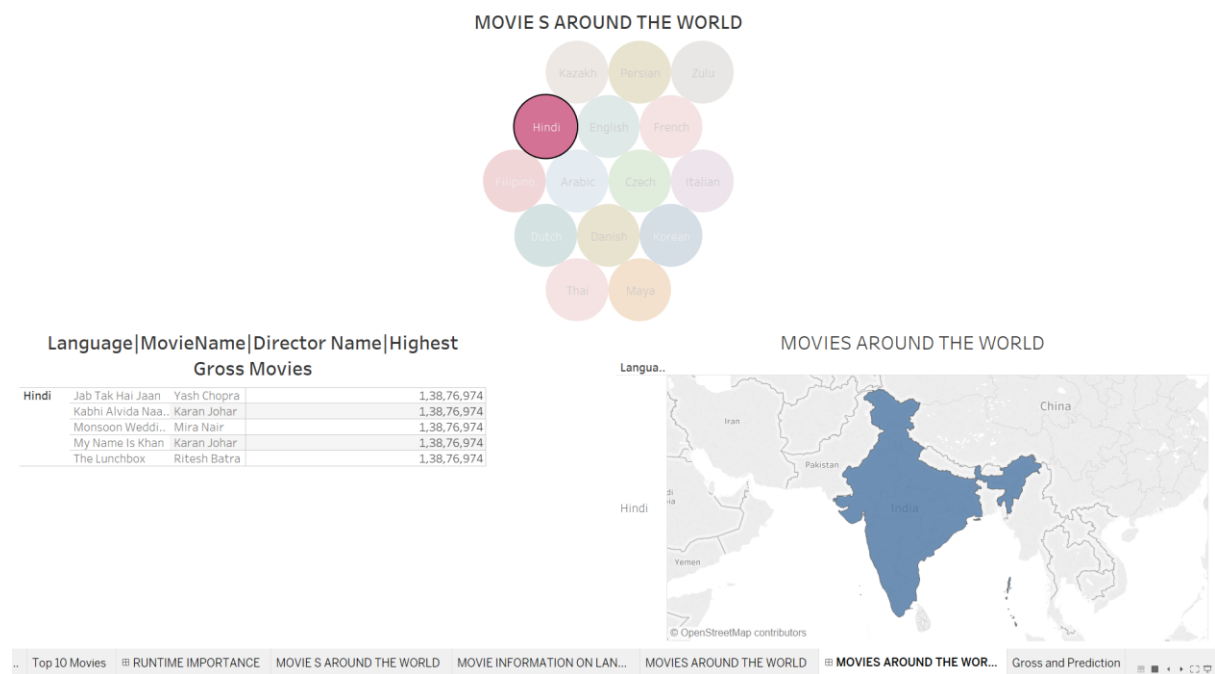
Audiences who enjoy movies around the world in every language can make use of this dashboard. This dashboard displays all the language movies around the world. Even it allows them to explore worldwide movies.

### Question:

List me all the movies based on language?

### INSIGHTS:

Our first slide displays the movies on all languages released till now. For now, we have filtered the top most profited movies on different languages are released. Second slide, displays the all the information of the movies like director name, movie name and gross made by that movie. Third slide displays the world map, to display which language does it belong to the country. Which will be helpful for all the film industry people as well as audiences.



## MOVIES AROUND THE WORLD



Language|MovieName|Director Name|Highest Gross  
Movies

Aboriginal	Rabbit-Proof Fence	Phillip Noyce	7,25,15,360
	The Interpreter	Sydney Pollack	7,25,15,360
Arabic	Caramel	Nadine Labaki	10,60,591
Aramaic	The Passion of Christ	Mel Gibson	4,99,263
Bosnian	In the Land of Blood and Honey	Angelina Jolie Pitt	3,01,305
Cantonese	2046	Kar-Wai Wong	3,23,33,860
	Exiled	Johnnie To	3,23,33,860
	Ip Man 3	Wilson Yip	3,23,33,860
	Kung Fu Hustle	Stephen Chow	3,23,33,860
	Rumble in the Bronx	Stanley Tong	3,23,33,860
	Shaolin Soccer	Stephen Chow	3,23,33,860
	The Legend of Drunken Master	Chia-Liang Liu	3,23,33,860
Czech	I Served the King	Jiri Menzel	6,17,228
Danish	Flame and Citron	Ole Christian Moe	16,47,780
	The Celebration	Thomas Vinterberg	16,47,780
	The Hunt	Thomas Vinterberg	16,47,780

## MOVIES AROUND THE WORLD



## 11<sup>th</sup> Dashboard:

### Target Audience:

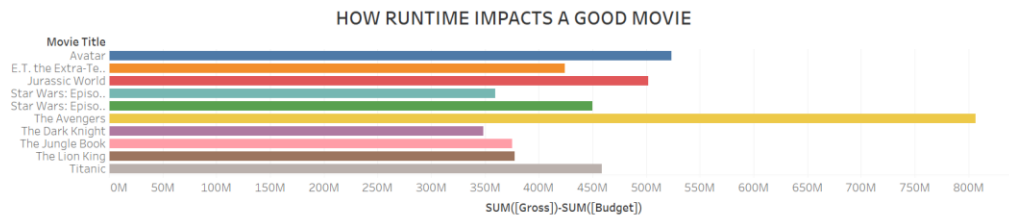
Runtime is more important to have a good movie. If a movie is too short, audiences will not be satisfied with the small movies. If the runtime is too long, **directors** have to be careful not to push the audiences in the bored stage. This will help the directors to choose the runtime on different genres.

### Question:

How much runtime should I have to keep in my movie?

### INSIGHTS:

Our slide displays most successful movies around the world. So, we collected all the movies based on profit. So that we can able to view the runtime used by them. All successful movies have the range from 2-2.30 hours. So that all the audiences have the satisfied feeling over the money spent on each movie. Displaying the information along with the language helps them to predict how much runtime should they keep based on languages. Directors name too mentioned, will help them to track based on other movies on these directors.



**SUCCESSFUL DIRECTORS AND MOVIE RUNTIME**

Director Name	Language	Movie Runtime	Profit
Steven Spielberg	English	185	2,486,332,231
George Lucas	English	142	1,386,641,480
James Cameron	English	194	1,199,625,910
Joss Whedon	English	173	1,000,886,628
Chris Columbus	English	174	941,707,624
Tim Burton	English	127	824,275,480
Christopher Nolan	English	169	808,227,576
Peter Jackson	English	201	777,968,050
Jon Favreau	English	126	769,381,547
Francis Lawrence	English	146	755,501,971
Michael Bay	English	184	644,242,537

FACEBOOK LIKES ON GENRES | GENRES CHOSEN BY ACTO... | Information About Runtime and ... | Top 10 Movies | **RUNTIME IMPORTANCE** | MOVIE S AROUND THE WORLD | MOVIE INFORMATION

## 12<sup>th</sup> Dashboard:

### Target Audience:

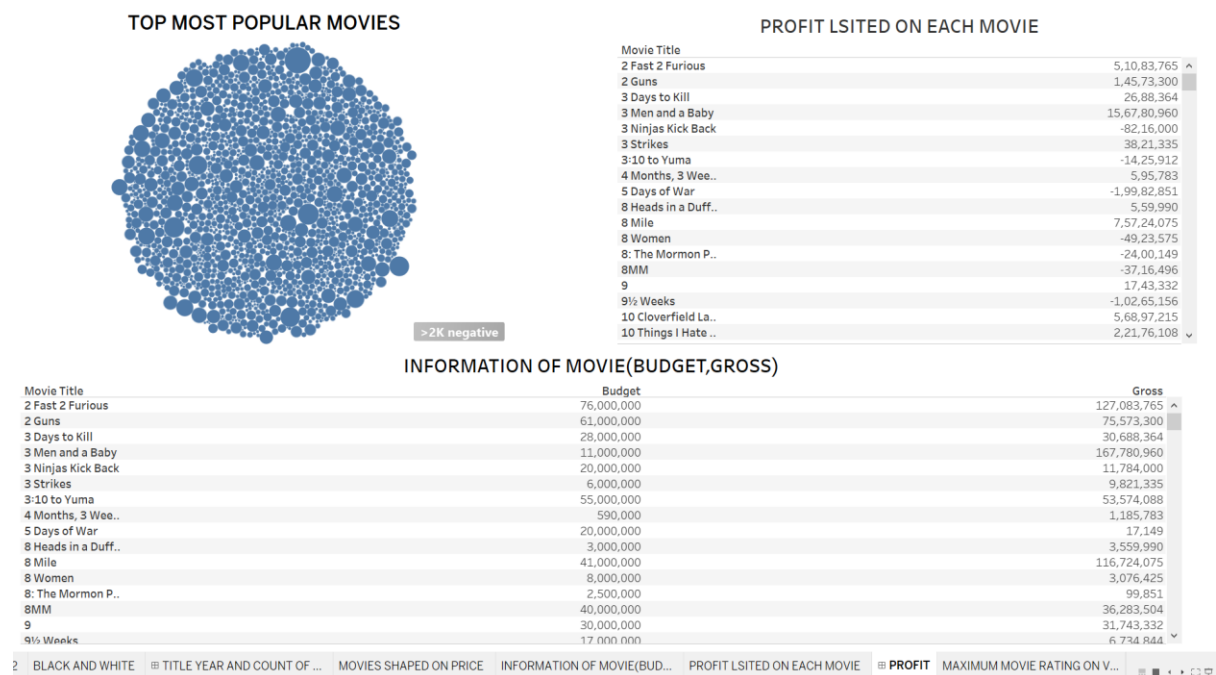
**Audiences** worldwide can make use of this dashboard. Because this dashboard consists of all overall information in the movies. Displaying the genres, budget, gross and profit made by the movies.

### Question:

Can I have the budget, gross and profit of entire movies?

### INSIGHTS:

This dataset contains the whole information of budget, gross and profit made by all the movies in the dataset. This dataset will be very useful for the audiences around worldwide to give the prediction and as well as analysis on the dataset. Thus, we use the profits of all movies which will be useful for the entire people involved in the cine industry. Thus, it gives depth core of analytics of the budget, gross and we use the formula in Tableau to calculate the profit of all the movies.  $\text{Sum}([\text{gross}]) - \text{sum}([\text{budget}])$ .



## 13<sup>th</sup> Dashboard:

### Target Audience:

**Film crew** will be utilising this dashboard. Because this dashboard displaying how many faces in the poster of all the movies. So we took the count of faces in number of poster.

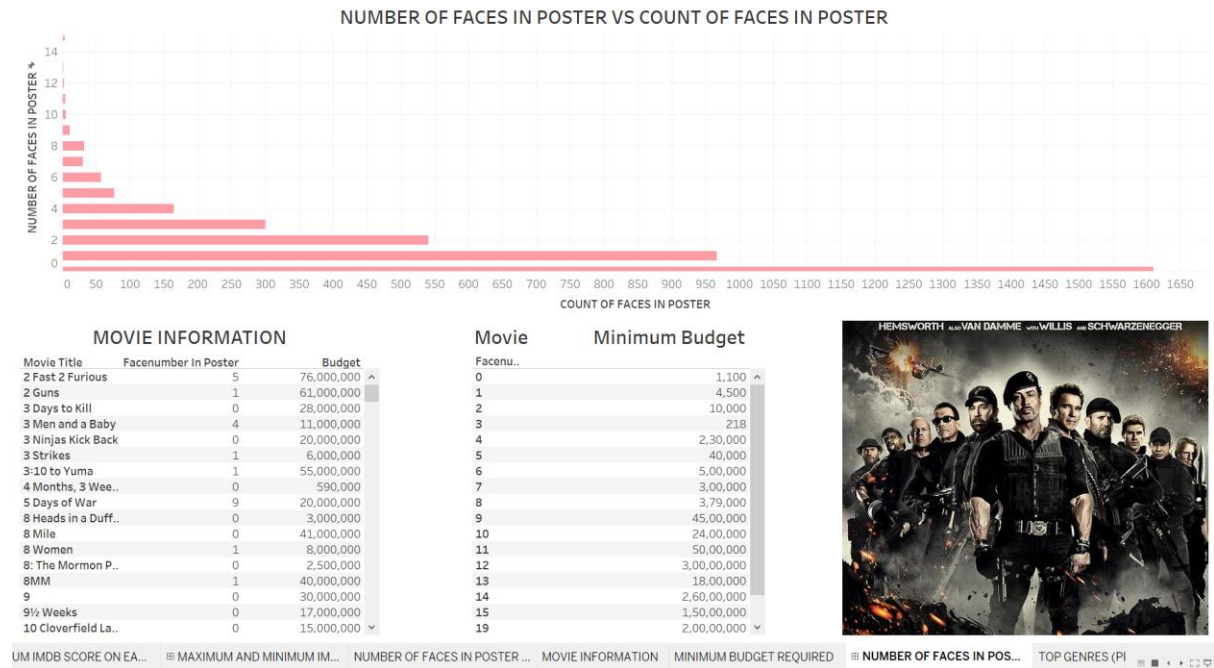
### Question:

How many faces should I put in my upcoming movie poster?

### INSIGHTS:

This dataset depicts the count of faces in posters and displayed. We can see how a poster is very important for a movie. It should display the partial contents of the movie as well as should display which actor have to be posted on the poster. Moreover, The more the number of actors in the movie. The more fans will watch the movie. One of the examples is the movie expendables. In this movie there are more than 10+ actors acted in the movie. Sylvester Stallone directed the movie and acted as a lead in this movie. So, he knows how to fetch the audiences. He is the first person to introduce a greater number of faces in a single poster. Which was encouraged worldwide regarding this poster and got a huge trademark to this movie.





## 14<sup>th</sup> Dashboard:

### Target Audience:

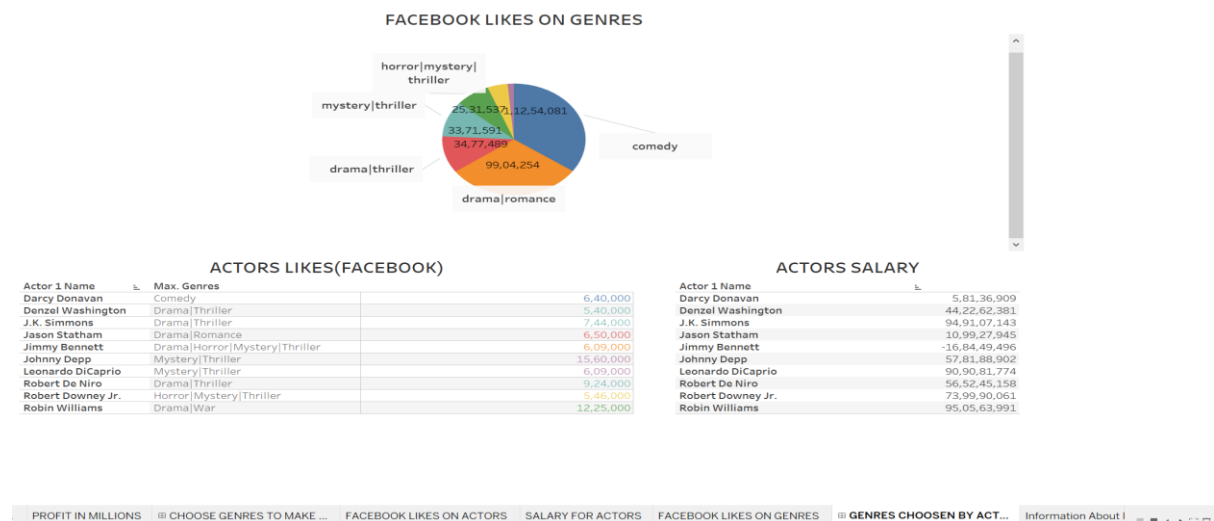
Likes on Facebook is very important to make a movie popular. Because the more likes it get, the more views on that movies. This dashboard will be helpful for the entire crew to give a best movie for the audiences.

### Question:

How to make my movie popular in Facebook?

### INSIGHTS:

Actors have the major impact on making the movie popular, for instance if a famous actor like Johnny Depp act in a movie. It automatically gets popular. So, Facebook is a good platform for making the movie reach over worldwide.



## 15<sup>th</sup> Dashboard:

### Target Audience:

Movie viewers are the ultimate target for this dashboard. We are looking at the topic of user reviews vs critic reviews. There is a lot difference between them. If someone follows reviewing a movie as a job. He is a critic reviewer and a normal user reviews a movie, is just a review.

### Question:

How reviews are important to movies?

### INSIGHTS:

Reviews are the important part of the movies. A good review might make a movie a huge gross around the worldwide. Even if a movie I good, and the reviews are bad. It led to great loss in money. To improve the reviews, a movie maker should know to make good Facebook likes and make his movie popular. We explained how to gain the Facebook likes in the previous slide. Every people like to voice out their opinion on movies they watched. Good movies will always get the good review and attracts more audiences and leads to good profit in the movie industry. Thus, a critic is more important for movie to make it more profitable or making it a loss in money. Therefore, whole film crew have to do some analysis on reviews before releasing the movie.

