**IMDB Movies:**

QUESTIONS:

1. All the movies ever made by gross revenue and duration with rating impact.
2. No of movies each year.
3. Genre of movies by rating
4. Favourite actor by order.

Avg. Revenue

No of Movies

Avg. IMDB Rating

Avg. Size

* Descriptions:

Movie Database (IMDB) gain insights into various aspects of the movie industry, such as box office performance, genre popularity, actor/director influence, and critical reception.

The project involves collecting and cleaning the data, which includes information on movie titles, release dates, box office revenue, ratings, director names, actor names, actor Facebook likes, director Facebook likes, etc. The project can be expanded to include sentiment analysis of user reviews, box office success prediction, and regional differences in movie preferences. Overall, the IMDB Movie Analysis offers a comprehensive approach to analyzing the movie industry and can provide insights to a wide range of stakeholders.

In the given dataset there are 5024 rows and 28 columns.

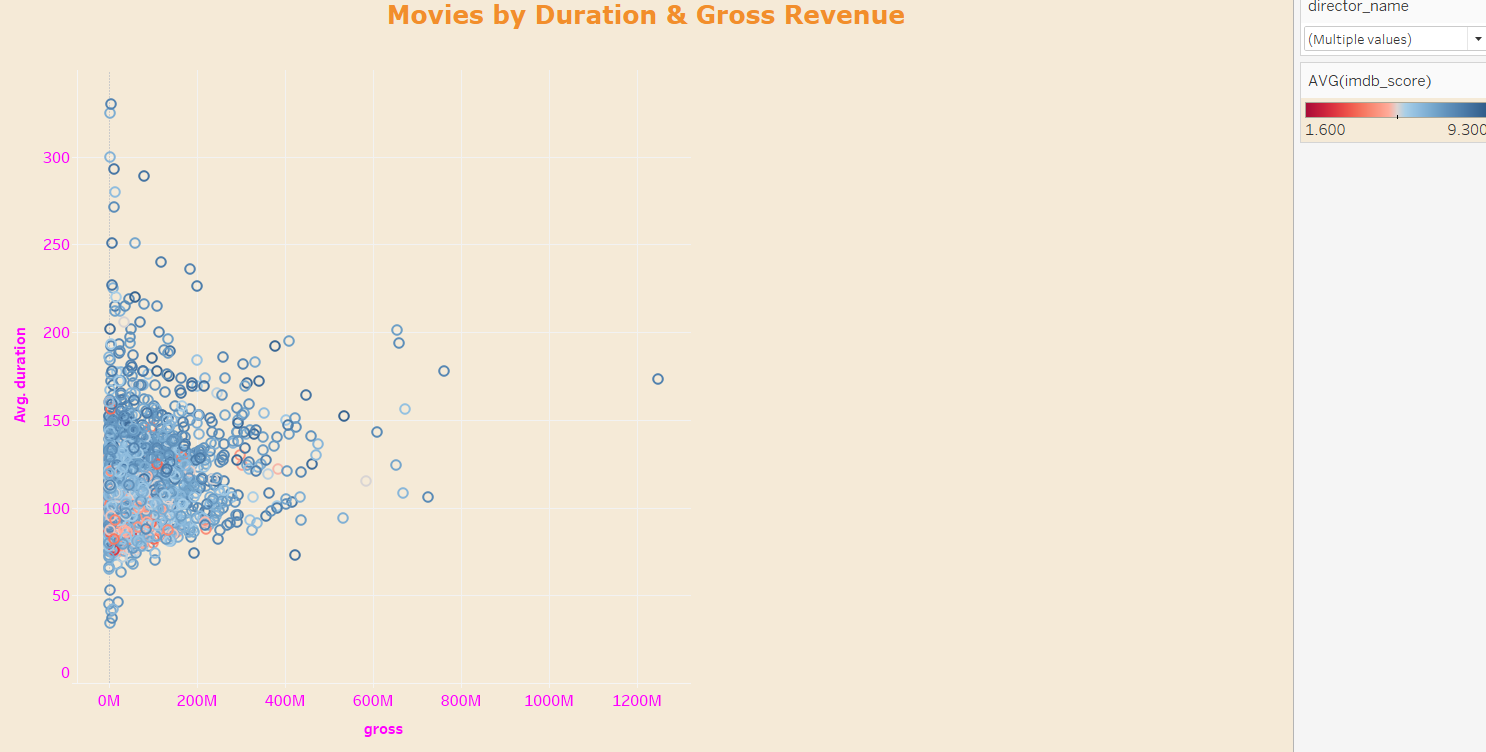
Tool: Tableau.

Cleaning the data in Tableau:

At the first stage in the tableau import the movies\_metadata.csv file. For cleaning from the “title\_year” column convert datatype from integer to date. After that, if we check the “genre” column there are many genres in one column. So for that column, we can split the genre or do custom splitting for all genres as we only get 3 splits of the genres.

* In the first worksheet, we are solving the problem of the total number of movies with an Average duration, Average IMDB Scores, and total gross(revenue) we are getting. But in the filter, I am using “director\_name” so we can see the total gross from the selected director who created movies and the industry gets the revenue. Or we can select all directors for total gross.

ANSWERS:

Qus-1: All the movies ever made by gross revenue and duration with rating impact.

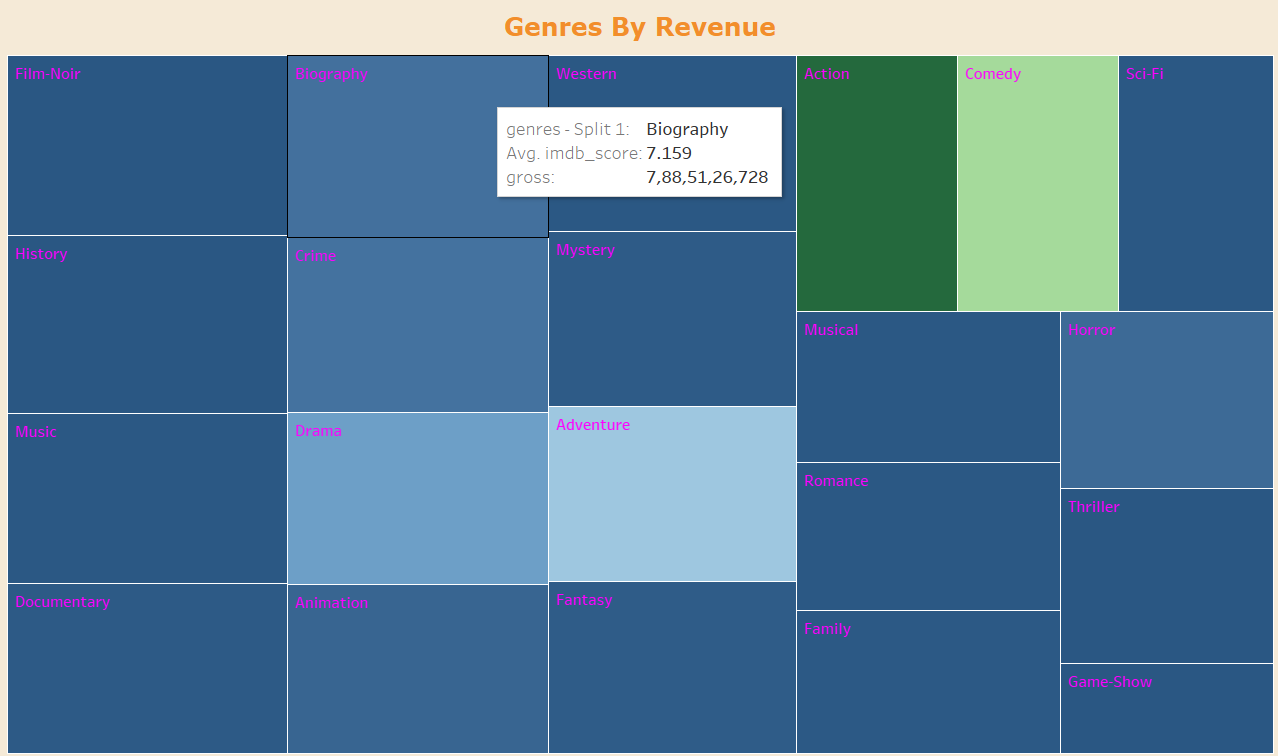
* Here, we want the relationship between revenue and the duration of the movies. So we are using a scatter plot for the correlation between gross and duration.
* Average IMDB\_score is low at that point we can see the orange colour and we can also see some outliers for that we can say that some movie’s average duration is high but also gross is high. People mostly prefer the average movie duration which gives the highest gross.

A picture containing text, diagram, line, plot

Description automatically generatedQus-2: No of movies each year.

* When we are talking about date and times we should use a line chart.
* At the first stage we already did the “title\_year” column’s datatype from integer to date so we get all years with the first day of every year.
* There is one big hike from the year 1990 to 2000 in the number of records.
* Director\_names added in the filter, so we also get the number of records of the movies which is created by the director each year.

Qus-3: Genre of movies by rating.



* For the genre we split the genre from the “Genre” column.
* Here I used a tree chart for the different genre’s ratings along with its revenue.
* From different genres industry gets the highest and lowest genre and also from the rating we get the customers’ genre choice.
* “film-noir” genre is the most preferable genre and the “Game Shows” genre is the least preferable genre as per the IMDB score rating.
* Also, here we added gross so the “Action” genre is getting the most revenue followed by the “comedy” genre. And “Music”, “Film-noir”, “History”, and “Sci-fi” genres are getting less revenue.

Qus-4: Favourite actor by order.

* Here, we are using a bar chart for the best actor throughout these years.
* Which actor is giving most of the revenue to the industry we can see by using this bar chart.
* I also include director\_name in the filter tool.
* So, from there we can check which actor and director worked together and give the most revenue to the industry.
* “Johhny Depp” is the most preferable actor by audience for the movies so industry getting most of the revenue by this actor.
* And talking about the director we can filter any director so we get the actor’s name which generated most of the revenue by working together.

A screenshot of a computer screen

Description automatically generated with low confidence

Insights:

* I understand Box office performance, Audience preferences, Critical reception, and Actor and director influence from the IMDB Data Set.

Result From the given data set I have observed the following result.

* Correlation between movie ratings and box office revenue.
* Comparison of ratings for different genres, directors, or actors.
* Identification of factors that influence movie ratings, such as runtime, or budget.
* Identification of the most influential actors or directors based on their involvement in highly rated movies.