Statistical Data Exploration: (EDA)

IMDB Movies Rating Data Analysis

Sub: IMDB Movies Rating

Tool: Tableau

QUESTIONS:

1. Is there any relationship between movie budget and revenue?
2. What are the duration outliers in various genres of movies?
3. How is the distribution of various movie duration?
4. Does having more Facebook likes have an impact on revenue?
5. Is there any relationship between various Facebook likes and IMDB voting?
6. Correlation matrix between various numerical data points
7. Is the genre budget changing as time is changing?
8. What is the distribution of IMDB ratings among various genres?
9. What is the most revenue-fetching category for a movie?
10. Facebook ratings relationship with movie ratings

Answers:

**QUS-1. Is there any relationship between movie budget and revenue?**

* We are using a scatter plot to show the relationship between Budget and revenue.
* Here, the relationship between Budget and revenue can be a positive correlation or sometimes it will negative correlation.

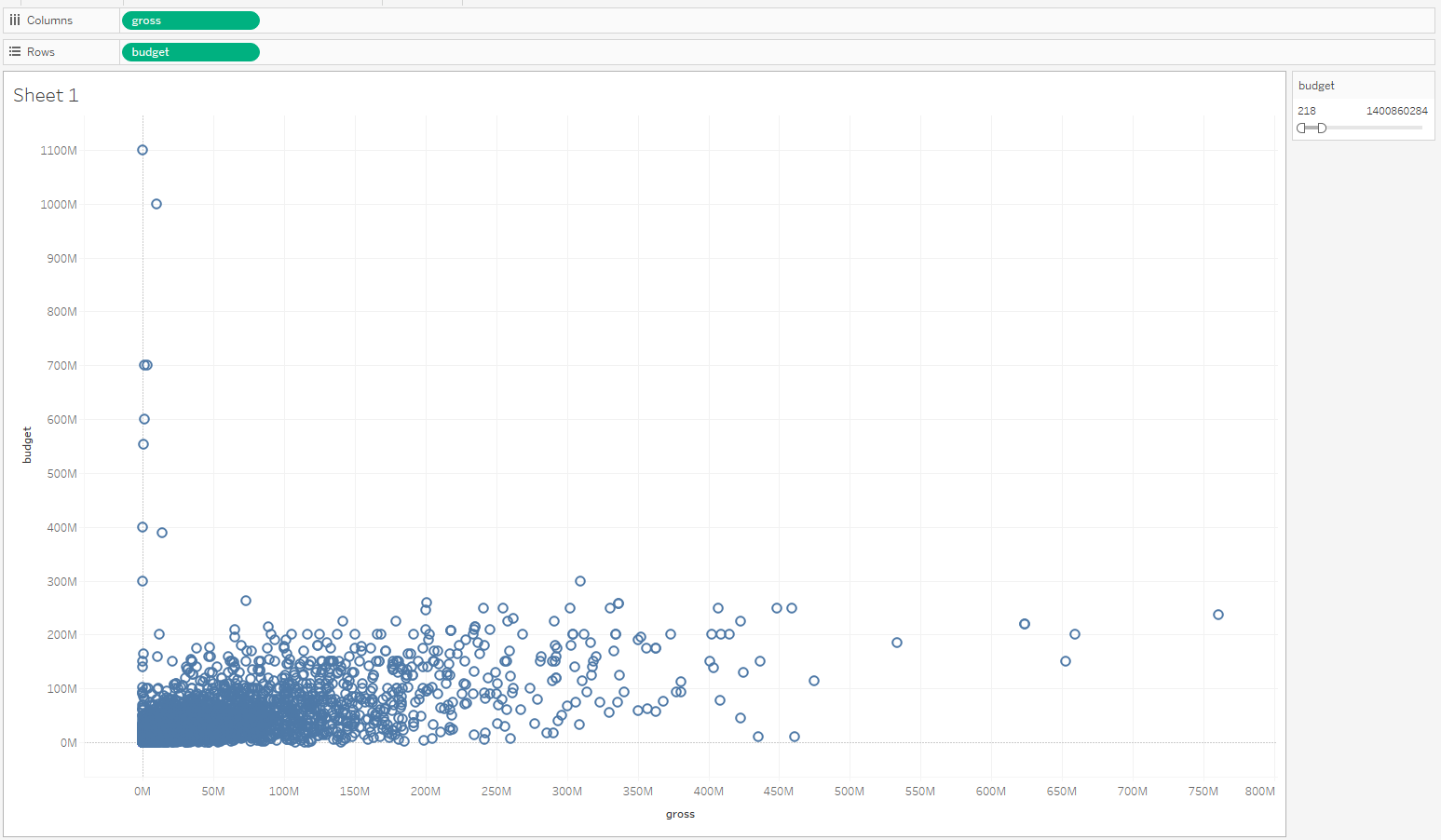
Here, we are filtering the budget. If we are going with a small-budget film correlation between budget and revenue gives a positive relationship, and we will not narrow down the budget - we can say if the budget is, more than 10M or above it will give a negative correlation between budget and revenue.

In the Tableau tool, we can filter out the budget according to its demand like from

218 -10M or 218 – 7M, 218 – 3M, etc.

From filtering, we can measure the relationship between Budget and revenue.







Now, let’s check with another filter. We will change the maximum budget from 218 - 11,82,77,06,356.

A screen shot of a graph

Description automatically generated with medium confidence



**Qus-2. What are the duration outliers in various genres of movies?**

* Here, we are using box plots for the outliers for the genres of the movies**.**

we already have the splits of the genres of the movies. I used split-1 for these outliers.

A picture containing text, screenshot, diagram, parallel

Description automatically generated

As we see in the box plot ‘Drama’ genre has the most outliers followed by Adventure, biography, Action, and crime comedy... but there are other genres that do not have outliers like Family, Music, Musical, Romance, Thriller, Western.

**Qus-3. How is the distribution of various movie duration?**

* A picture containing text, plot, screenshot, diagram

  Description automatically generatedfor the distribution of the movie duration we will use a Box plot.

Here, most of the films are created in nearly 100 minutes durations.

* *Generally, we are using box plot when questions are asked by interviewees related to distributions of movie duration or maybe sales duration distribution, etc.*

A screen shot of a graph

Description automatically generated with low confidence**Qus-4. Does having more Facebook likes to have an impact on revenue?**

* for the relationship between revenue(gross) and movies-Facebook likes, we are using a scatter plot.
* As we can see there is a positive correlation between revenue and Movies Facebook likes. That means if the Movies FB likes are high revenue will be also high.
* *For the relationship between two things like revenue and FB like or maybe in sales revenue and total sales monthly, yearly, etc. we will use a Scatter plot for the correlation. Either it is a positive – correlation or a negative correlation.*

**Qus-5. Is there any relationship between various Facebook likes and IMDB voting?**

* As we saw before many charts for the relationship between two values, we are using a Scatter plot.
* Here, also we are using a Scatter plot for the relationship between IMDB score and movies’ FB likes.
* When we checked in the Tableau tool for these two relationships, we conclude that if when IMDB scores are going above 5 it will create a positive relationship with FB Likes.
* We can say that it is not a linear relationship but a polynomial (line) shape we can see in the chart.
* For this relationship we can say that people are rated the same when IMDB scores are higher than 5.

**A picture containing text, diagram, screenshot, line

Description automatically generated**

**Qus-6. Correlation matrix between various numerical data points**

* Again, the correlation between two variables so a Scatter plot will be used.
* But in this question, we want a correlation matrix means we are using a scatter plot with many variables. So many scatter plots only in one chart and we can see if the correlation between two variables is positive or not.
* So far, we only use gross(revenue), duration, IMDB score, Movies FB likes.
* As we checked in the chart relation between budget and Budget gives a linear positive correlation, likewise gross with the gross, duration with duration, etc. gives a positive linear line. Gross with Movies\_FB\_like positive correlation, IMDB\_Score & Gross positive correlation, etc. relationship we can see only in the chart which is called correlation matrix with various variables.

A picture containing text, screenshot, line, diagram

Description automatically generated

**Qus-7. Is the genre budget changing as time is changing?**

* There are times and budgets in question.
* For the time we are using the title\_year column but for that, we are changing the year from the day so that we get the years that are available in the dataset.
* In this dataset we have 1962 to 2016 movies.
* We include the budget during these years with the genre.
* For that we are using a bar chart.
* As we see in the bar chart we get the different genres with years, where we conclude their constant change in the ‘Action’ genre as times change.
* A picture containing text, screenshot, plot, line

  Description automatically generatedBut there is one big change in the year 2006 when a sudden change in the genre of ‘Comedy’ went high as there are many ‘Comedy’ movies realized at that time but after that, it went down.

**Qus-8: What is the distribution of IMDB ratings among various genres?**

* as we already saw before chart for the distribution. We used a bar plot for it.
* Likewise, here again, we will use a bar chart for the distribution of IMDB ratings along with various genres.
* When we checked in the tableau tool for ratings and genres, we conclude that there is a left skew in the bar chart. So, we can say that people are ratings high for 5.5 to 7.5.
* And when we include various genres, we see that rating for the comedy genre is going continuously in the IMDB ratings, followed by the action and drama genres.
* Or if we want a specific genre for the rating for that first we should remove genre-Split 1 from the color bar and add genre-Split 1 into the filter bar. From there we can choose any genre which comes up with IMDB’s ratings.

**A picture containing diagram, text, line, screenshot

Description automatically generated**

**Qus- 9: What is the most revenue-fetching category for a movie?**

* In this question ask for the most revenue generated from which categories during these years.
* So we are using content\_rating for the category of movies with gross(revenue) along with years.
* Using “title\_year” we can exactly see the revenue of each year and how it is growing yearly or shrinking the revenue.
* As per our chart “PG-13” is a continuously growing category that is generating the highest revenue for a movie.
* But the “R” category of rating\_content is also the reason for the high revenue followed by “PG”, and “G”.
* These are the categories of “rating\_content” which generated most of the revenue throughout these years.

A picture containing text, screenshot, plot, line

Description automatically generated

**Qus- 10: Facebook ratings relationship with movie ratings.**

* We are using a scatter plot here to showcase relationship between IMDB\_score and Facebook likes.
* As we see in the chart after 5.0 ratings Facebook likes are going high.
* This is likes and scores are going high at the stage between 5.0 to 8.5.
* After seeing this chart we can say that there is positive correlation between IMDB\_score & FB\_Likes.

A picture containing text, screenshot, line, plot

Description automatically generated