**Movie project Analysis**

**Q1) Do movies become profitable with longer runtime?**

Analyzing movie runtime effect on profit%, Box office collections and number of releases. Here we can see that 90-115 runtime range has the highest box office collection and maximum no. of releases.

We can say that box office collections are not highly correlated to runtime range. In correlation coefficient test, we found that they have low positive correlation i.e., is 0.3.

Profit % bar graph is showing maximum profit in 160-185 runtime range and its increasing till 165-180 range which shows there is positive relation between them.

Robert Downey Jr is the most successful actor with 3 biggest blockbuster movies while Johnny Depp and Mark Hamill are on second position here with 2 movies.

Anthony Russo, Joe Russo are the top directors with 3 movies and Gore Verbinski comes second here with 2 biggest blockbusters.

Note- Biggest blockbuster is the movie which has the highest box office collection and profit in a year.

The income of the biggest blockbuster of the year is increasing most of the year and have their peak in 2010 and 2019. An exceptional drop can be seen in 2020 due to global pandemic.

**Q2) Are different genres more profitable than other genres?**

Chart, histogram

Description automatically generated

This graph shows us all the genres average income for each genre after averaging all movies in the data frame we have. So, from 2015-2020. This graph suggests that family movies would give the highest average gross income. So, it would suggest that making a family genre movie would give you a high income. The next best genres to make a movie judging from this table would be action and animation films. It also suggests that music and thriller genre movies produce less gross income on average.

Graphical user interface, text, application

Description automatically generatedGraphical user interface, application

Description automatically generated with medium confidence

However, when we look at the number of films in each genre then over the 20-year time we have we notice that we have very few movies in some genres. Thus, it is hard to make a prediction with any confidence about these genres with the limited data we have on them. Which is why I eliminated movies with under a certain number of movies that being 50. Then I display the new table to show the genres which have enough data to make some accurate predictions.

Notice are top average earner is the family genre is no longer in the table so I cannot suggest making a family movie as while it had highest average it doesn’t have enough data to support that conclusion. Also are lowest two earning genres thriller and music are no longer in the table. So, it isn’t certain that they would all be low average earners, but it is still not advisable to go for them.

Chart

Description automatically generated

Now displaying the new graph, we can see that our two top average earning genres with enough data to support that these averages are accurate are the Action and Animation genre. So, if you make a movie with that as the main genre then on average you will get a higher gross income. We can see that Crime is our lowest average earner so it’s probably best not to make a movie with that as the main genre. However, the other genres aren’t much better except for our two top earners and the adventure genre.

Chart, scatter chart

Description automatically generated

Examining the animation genre over the most recent five years in the data 2015-2020 and plotting its gross income on the y axis while date is on the x axis, we get the above graph. As you can see the graph has an even distribution with no peaks or troths. Which suggests that the animation movie genre is a stable earner and isn’t losing or gaining in popularity. Thus, it is more likely to make a consistent amount and from previous we know that animation has a high average earning thus it would be a good choice make movie with animation as the main genre.

Chart, scatter chart

Description automatically generated

Analysing our next highest earning genre over the past five years we see that it two has a consistent spread no matter what the year is meaning it too is a consistent earning genre. However, there are a few more points at the lower end of the gross income which makes it a bit riskier than the animation genre. However, it does have a higher on average income though I would say half of the movies in the last five year are low earners.

Chart, line chart, scatter chart

Description automatically generated

Looking at the movie gross income against year we see that each year is steady income wise however 2020 has few movies and they are low earners which may be due to pandemic.

Chart, bar chart

Description automatically generated

Now this is analysing films with main genre of animation between 2015-2020 while looking at the average gross income depending on what the sub genres are. Looking at this we can see that animation movies which have action, crime as their sub genres do terribly compared to the rest and make very little on average so it would not be advisable to make animation movie with those sub genres. It would also have brought down the values on our previous graphs. Looking at this I can recommend making an animation film with the sub genres adventure, family as that on average makes a lot. The other three bars in this graph also have some decent averages at least compared to the lowest and it wouldn’t be terrible to make an animation movie with those sub genres.

In conclusion making animation movie with the sub genres of adventure, family as the genre is the overall highest earning genre and the adventure, family are the two highest earning sub genres in our highest earning genre. In fact, looking at the previous graph of average gross income by genre we see that the highest three sub genres in this graph have a higher average gross income than the overall average for the entire genre in the previous graph looking at main genre only. So, action, adventure would be good to make as well as adventure, comedy. Looking at comedy family it is pretty much the game as average of the genre so overall would also be an okay choice.

Chart, histogram

Description automatically generated

Looking at the action genre and its sub genres for the past five years and their average gross income gives us the table above. Looking at this we see that Adventure, Sci fi sub-genre combo gives us the highest average gross income. In fact, looking at it and comparing it to previous tables we see that it is much higher than the average for the whole genre or even the animation genre. Which suggests it is a great combination choice to make a movie and that there are many sub-genre combinations that are pulling down the average and we can see that is true by looking at the graph above. In fact, many sub genres are underperforming in this genre compared to the average for the genre. For example, sub-genre horror, thriller or comedy drama or mystery sci-fi and would be poor choices to choose for making a movie.

Besides the highest average earning sub-genre of adventure, sci-fi the next three highest bars in this chart are also higher than the overall genre average. Meaning the next three highest would also be a good choice for this genre. So sub-genre of adventure, comedy would still be a good earning movie to make on average. Sub-genre of drama sci-fi is close to the average of the genre as a whole so would be ok movie to make.

Chart, histogram

Description automatically generated

Now we look at the average for all movies and their sub-genres for the past five years to see if there are any other surprises with genres with lower averages overall having a genre sub-genre which performs much higher than the overall average. We see that the Biography with the sub-genres drama, musical makes a much more on average than what would be suggested by its overall gross income for the genre. Drama with sub horror, sci-fi does much better than the average for its genre. So, both would be good choice for making movies in those genres. We also see several genre sub-genre combinations that don’t perform well like Drama with sub fantasy, Romance.

Chart, scatter chart

Description automatically generated

Now looking at something now to do with the genres. This is a graph of budget compared to gross income of the movie. As you can see there is a correlation between a budget and income with the higher the budget the higher the gross income is. However, this relationship is truer for lower budget movies meaning that you can predict the gross income a movie will make the lower the budget is. So, if you want to play it safe you make a lower budget movie however if you go for a higher budget movie there is much more of a chance that your movie will make much more or less than predicted by the line. Which means higher budget movies can have a greater payoff but are much riskier.

**Q3) Does the runtime of the movie affect the public rating?**

The figure shows that from 2000 to 2004 the runtime for movie was between 80 and 110 minutes with the average vote between 4.8 and 7.6. few movies had a runtime outside that range and their average vote is slightly over 6

As it can be observed in the other three figures, the data concentration changes for both runtime and average vote. From 2005 to 2009, for instance, the data cloud is more concentrated between 80 and 140 minutes for the runtime with the average vote between 5 and 7.4. It is also possible to observe that for this range of year more movies had a long runtime with relatively high average vote

The Dynamic is the same for the yearly range 2010-2014 and 2015-2020. In which the visualizations show different concentration of data.

**Q4 and Q5)**

**Does the genre of a movie determine or affects its public rating and runtime?**

**Determine the worldwide gross incomes per genre in the twenty years’ time frame and analyze them in relation to the average vote, average runtime, and fan reviews**

Linear regression line between the Average Runtime and average vote of each genre from 2000 t0 2020 has a positive bias. However, with an R-value equal to 0.354, the interdependence between the two factors is partially linear.

From the worldwide gross income bar chart, Action movie has the highest worldwide gross income with more than 91 billion dollars followed by animation with 49 billion dollars and comedy with 41 billion dollars, while as music has the lowest worldwide gross income with only a worldwide gross income of almost 8 million dollars

Regarding the average runtime per genre in the considerate time frame, Biography has the highest average runtime of 122 and thriller with runtime of 88. The Average runtime for the other genres, however, appears relatively uniform with no significant difference. In relation to the worldwide gross income, the regression line is almost flat showing an insignificant relation between the 2 factors

Concerning the average vote, the only difference with the previous analysis is that the regression line is negatively inclined, which highlights an inversely proportional relationship between the worldwide gross income and the average vote. Furthermore, the R-value is negative confirming the finding.

Regarding fan reviews, the data shows that in addition to the action genre which has the highest number of reviews in the 200-2020 range, the comedy genre ranks second followed by the drama genre in third position. this positioning differs from that obtained with the worldwide gross income by genres. In fact, in this ranking, the animation genre has a lower total number of fan reviews than comedy and drama. The data also show an almost perfect but non-linear relationship between the worldwide gross income by genre and the sum of fan reviews by movie genre. The linear correlation coefficient is equal to 0.9, a sign of a high linear interdependence

Statistical analysis was conducted to determine the relationship between the average vote and the average runtime for genres in four ranges of year. Therefore, the different linear correlation coefficients can be interpreted accordingly:

In the range 2000-2004, the regression line is decreasing, and the R-value is -0,133 which highlights the negative correlation between the average vote and the runtime. It can be observed that the data are far from the regression line.

On the contrary, the three other ranges of year have a positive regression line showing a positive dependence between the two values of the analysis as shown by the respective R-values. it is also shown that the intervals 2010-2014 and 2015-2020 have an R-value almost equal to 1 which shows an almost perfect linear correlation between the average vote and the average runtime.

The range 2015-2020 has an R-value equal to 0.994 which shows the highest dependence of the average vote on the average runtime

**Q6) are there directors who are more favorites, how good are the films they have directed so far and their career in general?**

By trying to look at the Movie Directors that are more favoured from the data frame containing a total of 955 Directors. Three directors were found to be more favoured by looking at directors that has produced more than seven (7) movies. Clint Eastwood directed 13 movies, Steven Soderbergh directed 11 movies and David Gordon Green directed 8 movies.

While three are the most favoured and have gathered a total of 34 movies among them between 2000 and 2020, they have contributed a total of about 1.62% worldwide gross income among them, with Clint Eastwood 0.8%, Steven Soderbergh 0.74%, and David Gordon Green 0.08%. This is a significant amount compared to the rest of 952 directors contributing the remaining approximately 98.4%.

Of the three directors, Clint Eastwood has made 173.29% profit percentage, Steven Soderbergh 260.36% profit percentage and David Gordon Green 10.73% profit percentage. While Clint Eastwood has been more favoured in terms of numbers of movies directed, Steven Soderbergh has made higher returns (profit percentage) per movies directed using the difference between average budget and worldwide gross income. Using a pie-chart to compare all three directors, David Gordon Green is the lowest at 2.41%, followed by, Clint Eastwood at 39.00% and Steven Soderbergh at 58.59%. With this, Steven Soderbergh is the more successful in terms of return on investment.

Looking at the average rating (vote), the boxplot shows that Clint Eastwood has better average rating for movies he has directed at 7, compared to the other two. However, the mean rating being above the median for Steven Soderbergh average rating, shows that Steven Soderbergh has more frequency of high rating than low rating in the movies he has directed.

From the line-chart, Clint Eastwood has had an average career in terms of worldwide gross income, with his most successful movie being is 2015 hit (American Snipper) an action movie. He is usually more successful in Action Genres and not as successful in Biography Genre, although he has produced more Biography Genre.

Steven Soderbergh usually have upward followed by a downward curve in terms of movies gross income. And that is because Steven Soderbergh usually follows an Action genre with a Drama genre, while the Action Genre are usually very successful, the Drama are not as successful. And this is indicated by the upward followed by a sharp download curve of his line-chart.

**Q7) Does the movie release time affect the profitability the movie?**

Looking at graphs we can say that May, December months are best for the release of movies with four biggest blockbusters more than twenty-five and twenty blockbusters and the lowest flops respectively.

July and October are not the best time to release movies since they produced around thirty-five and thirty-eight flops respectively and low no. of blockbusters.

**Hypothesis test**

Our assumption is that high fan reviews per genre is directly proportional to the worldwide gross income. In both Total Worldwide Gross Income and Total fan reviews per genre, action movie has the highest values when considering the period of our analysis. Therefore, we want to confirm (Hypothesis) or negate (Null Hypothesis) that the assumption is verified.

from the T-test, the P-Value is 0.039 which below 0.05 which lead us to reject the Null-Hypothesis and confirm the direct proportionality between profit and fan reviews.

even though the hypothesis is verified, the animation genre is the exception. In the Total worldwide Gross Income bar chart, Animation was the genre with the second best total gross Income. its position is different when considering the fan reviews. In fact, animation is at the fourth position after comedy and Drama which were respectively at the third and fourth position in terms of total worldwide gross income