

PROJECT 5:

IMDB MOVIE ANALYSIS:

DESCRIPTION:

The focus of this project is on figuring out what makes a movie successful on IMDB, where high ratings are a measure of success. The objective is to determine which aspects of a movie's success are most important by looking at a variety of factors like genre, cast, budget, and release date. When planning and making new films, this information can help movie producers, directors, and investors make better decisions. Understanding these factors may result in future films with greater commercial success.

APPROACH:

- **DATA GATHERING:**

I downloaded the data provided and Imported that data in the Excel workbook.

- **DATA CLEANING:**

The next step I followed is data cleaning. There were many blank cells in the file which needed to be removed to get the accurate output. Also, there were cells with not required data which needed to be removed.

- **INSIGHTS:**

The final step is to perform various operations including sorting, Mean, Median, Mode, Variance, etc. to derive outputs so that we can gather insights from the data.

TECH - STACK USED:

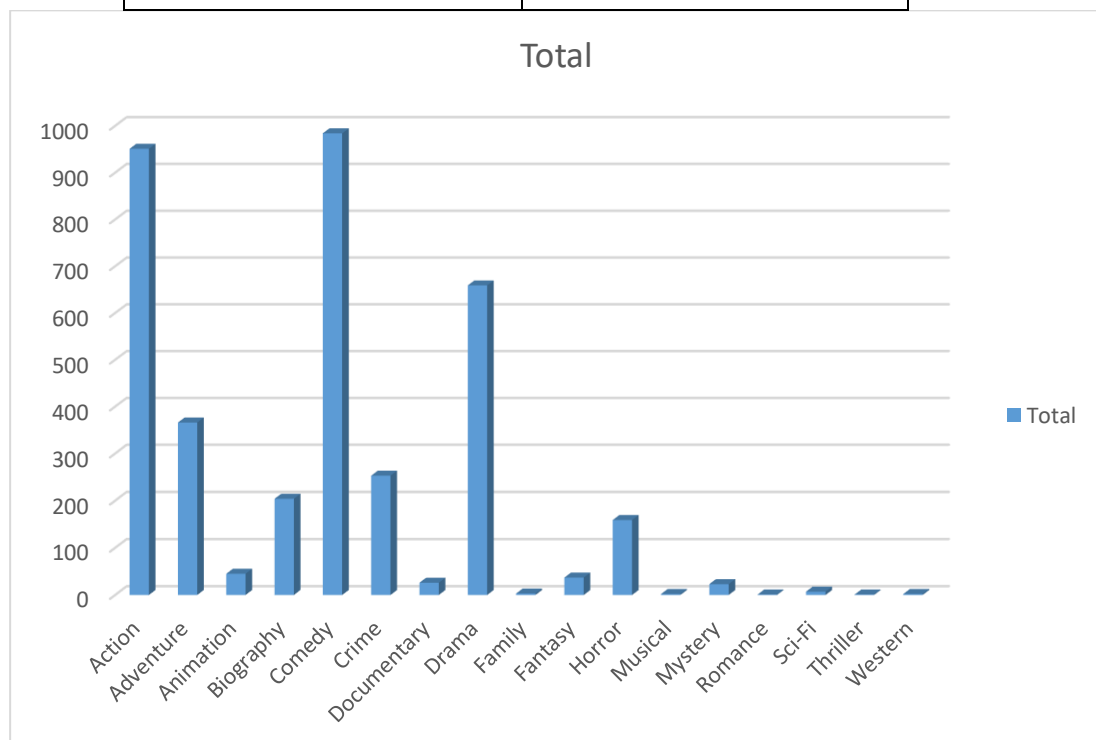
The Tech-stack used to complete this project is MS-EXCEL, 2013. The purpose of using MS-Excel is that it provides features like PIVOT TABLE, operations like MAX, MIN, SUM, COUNTIF etc. using which we can derive insights easily and data visualization is also possible using excel.

DATA ANALYTICS TASKS:

A. **Movie Genre Analysis:** Analyse the distribution of movie genres and their impact on the IMDB score.

Task: Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

| | |
|-----------|----------|
| Mean- | 219 |
| Median- | 37 |
| Mode- | 2 |
| Variance- | 103595.9 |
| STDEV- | 321.8631 |
| Max- | 984 |
| Min- | 1 |



INSIGHTS:

The most common Genres are:

Comedy, Action, Drama, Adventure and Crime.

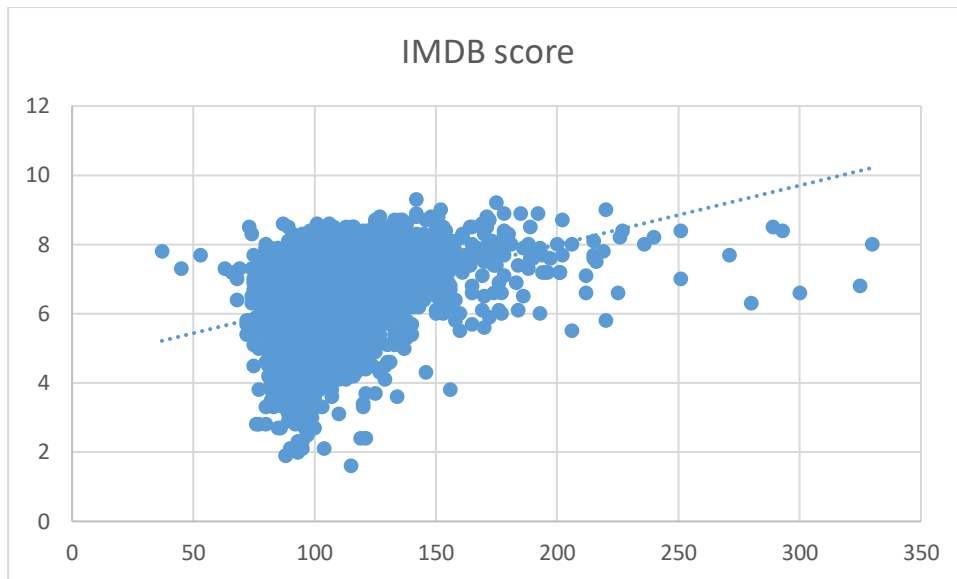
| Average of imdb score | Row Labels |
|--------------------------|-------------|
| 6.29 | Action |
| 6.56 | Adventure |
| 6.74 | Animation |
| 7.16 | Biography |
| 6.17 | Comedy |
| 6.94 | Crime |
| 6.80 | Documentary |
| 6.83 | Drama |
| 6.50 | Family |
| 6.28 | Fantasy |
| 5.84 | Horror |
| 6.75 | Musical |
| 6.65 | Mystery |
| 7.10 | Romance |
| 6.63 | Sci-Fi |
| 4.80 | Thriller |
| 8.10 | Western |

With the above data it can be inferred that Biographies, Romantic and Western movies are slightly liked more by the viewers.

B. Movie Duration Analysis: Analyse the distribution of movie durations and its impact on the IMDB score.

Task: Analyse the distribution of movie durations and identify the relationship between movie duration and IMDB score.

| | |
|---------|----------|
| Mean- | 110.2635 |
| Median- | 106 |
| stdev- | 22.67528 |



INSIGHTS:

There is a wide range of movie based on the duration. The duration can be as short as 37 minutes and as long as 330 minutes. As we can infer from the scatterplot that the movie with 37 minutes watch time can have an IMDB rating more than a movie with 89 minutes of watch time. With this we can say that duration can't alone impact the IMDB ratings. However, it is also seen that increase in duration has increased ratings. So it might lead to increase in ratings but it's not the deciding factor.

C. Language Analysis: Situation: Examine the distribution of movies based on their language.

Task: Determine the most common languages used in movies and analyse their impact on the IMDB score using descriptive statistics.

| Language | Count of movie_title | Average of imdb_score |
|------------|----------------------|-----------------------|
| English | 3566 | 6.43 |
| French | 34 | 7.36 |
| Spanish | 23 | 7.08 |
| Mandarin | 14 | 7.02 |
| Japanese | 10 | 7.66 |
| German | 10 | 7.77 |
| Italian | 7 | 7.19 |
| Cantonese | 7 | 7.34 |
| Portuguese | 5 | 7.76 |

INSIGHTS:

The Top five languages are English, French, Spanish, Mandarin and Japanese.

English being the most common language has the average imdb score less than the other 8 languages mentioned above. However, since the English movies are more in number, it can impact the average score. So, it can be inferred that the language has some personal favouritism for the viewers.

D. Director Analysis: Influence of directors on movie ratings.

Task: Identify the top directors based on their average IMDB score and analyse their contribution to the success of movies using percentile calculations.

| Row Labels | Average of imdb_score |
|-----------------|--------------------------|
| Akira Kurosawa | 8.70 |
| Tony Kaye | 8.60 |
| Charles Chaplin | 8.60 |
| Ron Fricke | 8.50 |
| Majid Majidi | 8.50 |

INSIGHTS:

Akira Kurosawa is the most favourite director followed by Tony Kaye and Charles Chaplin.

maximum= 8.7 Akira Kurosawa

TOP 5 Directors :

| | |
|-----------------|-----|
| Akira Kurosawa | 8.7 |
| Tony Kaye | 8.6 |
| Charles Chaplin | 8.6 |
| Ron Fricke | 8.5 |
| Majid Majidi | 8.5 |

| | |
|---------------|-----|
| Large= | 8.7 |
| Percentrank = | 1 |
| Percentile= | 8.7 |

From above we can infer that people like to watch their favourite director's movies and rate them highly as compared to others.

E. **Budget Analysis:** Explore the relationship between movie budgets and their financial success.

Task: Analyse the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

max (Profit margin)= 523505847
correlation coefficient= 0.0982896

| Profit Margin | Movie Title |
|---------------|------------------------------------|
| 523505847 | Avatar |
| 502177271 | Jurassic World |
| 458672302 | Titanic |
| 449935665 | Star Wars: Episode IV - A New Hope |
| 424449459 | E.T. the Extra-Terrestrial |

The Top Five movies with Highest Profit margin are mentioned above. Among these Avatar.

| Budget | Gross | Profit Margin | Movie Title |
|-------------|---------|---------------|-------------|
| 12215500000 | 2201412 | -12213298588 | The Host |

Movie with the maximum budget has recorded a loss.

From this we can infer that having a high budget of the film cannot ensure high profit.

RESULT:

From this Project we can draw an overall insight that there are certain genres, and directors that are liked by the audience more than other genres and directors and they prefer to watch movies in English. Duration of the movie cannot be the decisive factor in increasing the ratings of the movie. Also having a high budget and long duration cannot ensure high ratings, rather content and engagement of the viewers is more important.

[Imdbexcel](#)