##### **PROJECT:** **IMDB Movie Analysis**

**PROJECT DESCRIPTION:**

IMDB Movie Analysis Project is subjected to determining about the major underlying trends about the Movies Statistic Trends. In this Trends data statistic are reflected by taking mostly the Score obtained throughout the movie parameters. While looking for the analytics report, most of the parameters are subjected to Time duration, Budget of movie, Profit earned, Popular Genres with Score statistics for more refined statistical approach.

In this Project the Statistical comparison are subjected to the genes and type of movie, duration of movies reflecting to score pattern, most common language in which the movies are produced in, Director’s contribution to the success of a movie and the most important of them all is the Profit earned from those movies.

# APPROACH:

# Downloaded dataset attached in csv format and carried out further computation based on the questions and queries highlighted in this Project. As the requirement for technical computation without any calculus error, cleaning of data is a crucial part. Removing all the null or zero values from mostly the movie score, directors, Budget and related. Creating a copy of the sets of refined data and applied filters to manipulated the data and further used statistics functions like countif, sum, avg, stdev, median and related. Plotting of Graph by extracting the final records output in suitable fashion.

# TECH-STACK USED:

Used Microsoft Excel for Table Query, Computation, and Graphical Representation.

# INSIGHTS:

In this project, most of the statistics are curated by taking movies scores into consideration. As to strengthen the outcome of a result for certain tests, taking the scores is the foundation of the queries in countifs functions and related. In a contrast it could be shown that the duration of the movie is having the highest impact on the Score and likes obtained for the movies respectively.

# RESULTS:

While working on this project, I came to know that the category of this movies is very common throughout the statistics. As per a general part of approach it is proved perfectly that most of the block Buster movies or the seven-star rated movies are very epic and have the longest length of the screen time along with upcoming sequels. And those with the highest screen time / the duration are rated to be with the highest scores and surprisingly the average duration movies are awarded with average score rates which is quite amusing and informative.

Mentioned below are the Questions along with there Queries performed & Outputs received;

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

**OBJECTIVE:**  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.

------------------------------------------------------------------------------------------------------------------------------------------

# ANS:-

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **GENRE** | **MOVIES COUNT PER GENRE** | **IDBM SCORE STATISTICS** | | | | | | |
| **MEDIAN** | **AVERAGE** | **MODE** | **MAX** | **MIN** | **VAR** | **STDEV** |
| Drama | 236 | 6.90 | 6.76 | 7.20 | 9.30 | 2.00 | 0.92 | 0.96 |
| Comedy | 209 | 6.30 | 6.20 | 6.70 | 9.50 | 1.70 | 1.19 | 1.09 |
| Horror | 71 | 5.90 | 5.84 | 6.20 | 8.70 | 2.20 | 1.28 | 1.13 |
| Documentary | 51 | 7.40 | 7.18 | 7.50 | 8.60 | 1.60 | 1.12 | 1.06 |
| Thriller | 20 | 6.40 | 6.31 | 6.10 | 9.00 | 2.20 | 1.11 | 1.05 |
| Western | 12 | 6.80 | 6.69 | 6.50 | 8.90 | 3.80 | 1.09 | 1.04 |
| Action | 11 | 6.30 | 6.24 | 6.10 | 9.10 | 1.70 | 1.25 | 1.12 |
| Family | 5 | 6.40 | 6.25 | 6.70 | 8.70 | 1.70 | 1.44 | 1.20 |
| Romance | 4 | 6.50 | 6.45 | 6.50 | 8.60 | 2.10 | 0.99 | 1.00 |
| Adventure | 3 | 6.60 | 6.44 | 6.70 | 8.90 | 1.90 | 1.28 | 1.13 |
| Sci-Fi | 3 | 6.40 | 6.28 | 6.70 | 8.80 | 1.90 | 1.47 | 1.21 |
| Animation | 2 | 6.70 | 6.58 | 6.70 | 8.60 | 1.70 | 1.30 | 1.14 |
| Fantasy | 2 | 6.40 | 6.31 | 6.70 | 8.90 | 1.70 | 1.35 | 1.16 |
| Biography | 1 | 7.20 | 7.15 | 7.00 | 8.90 | 4.50 | 0.52 | 0.72 |
| History | 1 | 7.20 | 7.08 | 7.50 | 8.90 | 2.00 | 0.79 | 0.89 |
| Music | 1 | 6.70 | 6.46 | 7.10 | 8.50 | 1.60 | 1.44 | 1.20 |
| Musical | 1 | 6.70 | 6.51 | 7.00 | 8.50 | 2.10 | 1.50 | 1.23 |

# OUTPUT:- [Refer to this Link](PR-5%20A1.xlsx)

# Top 10 common Genres are “Drama”, “Comedy”, ” Horror”, ”Documentary”, ”Thriller”, ”Western”, ”Action”, ”Family”, ”Romance”, and “Adventure” a per the dataset given in the attachment. Refer the above-mentioned table for detailed statistical report.

1. **Movie Duration Analysis:** Analyze the distribution of movie durations and its impact on the IMDB score.

**OBJECTIVE**: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

------------------------------------------------------------------------------------------------------------------------------------------

# ANS:-

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **MOVIE DURATION ANALYSIS** | | | | | |
| **SR NO** | **DURATION RANGE** | **COUNT** | **AVERAGE SCORE** | **MEDIAN** | **STDEV** |
| 1 | 0-50 | 71 | 7.35 | 7.50 | 1.00 |
| 2 | 51-100 | 2108 | 6.04 | 6.20 | 1.20 |
| 3 | 101-150 | 2644 | 6.66 | 6.70 | 0.94 |
| 4 | 151-200 | 169 | 7.41 | 7.60 | 0.91 |
| 5 | 201-250 | 23 | 7.58 | 7.70 | 0.86 |
| 6 | 251-300 | 9 | 7.49 | 7.70 | 0.88 |
| 7 | 301-550 | 4 | 7.68 | 7.85 | 0.62 |

# OUTPUT: [Refer to this Link](PR-5%20A2.xlsx)

# Based on the data referring to the table mentioned above, All time favorite movies which are at an average score stat and duration of 151-200 are preferred.

# The Highest preferred and appreciated movie with generous score lies into the range of 301-500 duration category with the lowest deviation in the score range of 0.62.

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

**OBJECTIVE**: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

-------------------------------------------------------------------------------------------------------------------------------------

# ANS:-

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **LANGUAGE STATISTICS** | | | | |
| **LANGUAGES** | **COUNT** | **MEAN** | **MEDIAN** | **STDEV** |
| English | 4704 | 1.6 | 6.5 | 1.12 |
| French | 73 | 4.9 | 7.2 | 0.73 |
| Spanish | 40 | 4.4 | 7.15 | 0.86 |
| Hindi | 28 | 2.8 | 6.95 | 1.40 |
| Mandarin | 26 | 3.2 | 7.05 | 1.04 |
| **German** | **19** | **4.9** | **7.6** | **0.95** |
| **Japanese** | **18** | **5.6** | **7.6** | **0.99** |
| Cantonese | 11 | 5.3 | 7.2 | 0.70 |
| Italian | 11 | 5.1 | 7.3 | 1.24 |
| Russian | 11 | 4.1 | 6.5 | 1.38 |

# OUTPUT:- [Refer to this Link](PR-5%20A3.xlsx)

# Based on the statistics presented in the table above, “German” and “Japanese” are the 2 most common languages with the highest Score ratio of 7.6 and 7.6 respectively; whereas, “English”, “French”, and “Spanish” are the languages under which most movies are casted but lagged in Score Ratio with 1.1, 0.4, 0.45 points respectively.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

1. **Director Analysis:** Influence of directors on movie ratings.

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

------------------------------------------------------------------------------------------------------------------------------------------

**ANS:-**

|  |  |
| --- | --- |
| **DIRECTOR'S SUCCESS STATISITCS** | |
| **DIRECTOR'S NAME** | **DIRECTOR'S RATING** |
| John Blanchard | 9.5 |
| Frank Darabont | 9.3 |
| Francis Ford Coppola | 9.2 |
| John Stockwell | 9.1 |
| Christopher Nolan | 9 |
| Francis Ford Coppola | 9 |
| Peter Jackson | 8.9 |
| Steven Spielberg | 8.9 |
| Quentin Tarantino | 8.9 |
| Sidney Lumet | 8.9 |

# OUTPUT:- [Refer to this Link](PR-5%20A4.xlsx)

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

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

**OBJECTIVE**: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

-------------------------------------------------------------------------------------------------------------------------------------

**ANS:-**

|  |  |  |  |
| --- | --- | --- | --- |
| **MOVIE BUDGET STATISTICS** | | | |
| **MOVIE TITLE** | **BUDGET** | **GROSS EARNING** | **PROFIT EARNED** |
| **#** | **Million** | **Million** | **Million** |
| Star Wars: Episode IV - A New HopeÂ | 11 | 460 | 449 |
| E.T. the Extra-TerrestrialÂ | 11 | 434 | 424 |
| The Lion KingÂ | 45 | 422 | 377 |
| Star Wars: Episode I - The Phantom MenaceÂ | 115 | 474 | 359 |
| The Hunger GamesÂ | 78 | 407 | 329 |
| DeadpoolÂ | 58 | 363 | 305 |
| The Hunger Games: Catching FireÂ | 130 | 424 | 294 |

# OUTPUT:- [Refer to this Link](PR-5%20A5.xlsx)

# The Analytics visualized in the above-mentioned Graph is Plotted on the Basis of the movies produced with the “Lowest” Budget and Procured “Highest” Profit progressively.

**-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x---END-OF-DOCUMENT---x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-**