**IMDB Movies Analysis**

IMDB is well-known movie and series rating site for users and critics worldwide. In IMDB, we can find movies of series rating as well as its directors and actors compiled profile as well as financials of it. Here we are provided with IMDB dataset for movies from 1920 to 2010, which contains information about movies, its actors, directors, budget, collection, etc. We will be going to clean the dataset and get answers for asked questions by using 5 Why method for analytics using Office 365 Excel.

**Approach**

**In this data**set first we will understand the data then we will clean the data as per our requirement by removing duplicates and null values. Delete unnecessary columns.

We will be using pivots table, functions, charts to answer our various questions.

**Tech Stack Used**

We would be using MS Excel to clean and visualize data and MS word to create the reports with desired outputs.

Insights

1. **Cleaning the data**: This is important step as we need to clean our data before analysing it.

Your Task: Clean the data

1. Remove unnecessary columns.
2. Remove duplicate rows
3. Remove Blank Cell.
4. **Movie Genre Analysis:** Analyze the distribution of movie genres and their impact on the IMDB score.

|  |  |
| --- | --- |
| **Row Labels** | **Count of genres** |
| Action | 951 |
| Adventure | 366 |
| Animation | 45 |
| Biography | 204 |
| Comedy | 984 |
| Crime | 253 |
| Documentary | 25 |
| Drama | 659 |
| Family | 3 |
| Fantasy | 37 |
| Horror | 159 |
| Musical | 2 |
| Mystery | 23 |
| Romance | 1 |
| Sci-Fi | 7 |
| Thriller | 1 |
| Western | 2 |
| (blank) |  |
| **Grand Total** | **3722** |

|  |  |
| --- | --- |
| Average | 218.94 |
| Median | 37.00 |
| Mode | 2.00 |
| Max | 984.00 |
| Min | 1.00 |
| Variance | 110094.81 |
| Standard Devi. | 331.81 |

From the above table we can see that the highest movie genre is comedy followed by Action and minimum genre movie are romance, thriller.

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

To see the impact of duration on IMBD ratings we need to plot the scatter plot chart.

|  |  |
| --- | --- |
| **Average** | 110.265717 |
| **Median** | 106 |
| **St. Deviation** | 22.68 |

Graph:

Through the graph we can see that the duration of 60 to 150 gives the ideal range of IMBD score i.e. between 4 to 9 and as we increase the durations the IMDB score is max to 8.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | **Count of language** |  | **Row Labels** | **Average of imdb\_score** |
| Aboriginal | 2 |  | Aboriginal | 7.0 |
| Arabic | 1 |  | Arabic | 7.2 |
| Aramaic | 1 |  | Aramaic | 7.1 |
| Bosnian | 1 |  | Bosnian | 4.3 |
| Cantonese | 7 |  | Cantonese | 7.3 |
| Czech | 1 |  | Czech | 7.4 |
| Danish | 3 |  | Danish | 7.9 |
| Dari | 2 |  | Dari | 7.5 |
| Dutch | 3 |  | Dutch | 7.6 |
| English | 3566 |  | English | 6.4 |
| Filipino | 1 |  | Filipino | 6.7 |
| French | 34 |  | French | 7.4 |
| German | 10 |  | German | 7.8 |
| Hebrew | 1 |  | Hebrew | 8.0 |
| Hindi | 5 |  | Hindi | 7.2 |
| Hungarian | 1 |  | Hungarian | 7.1 |
| Indonesian | 2 |  | Indonesian | 7.9 |
| Italian | 7 |  | Italian | 7.1 |
| Japanese | 10 |  | Japanese | 7.5 |
| Kazakh | 1 |  | Kazakh | 6.0 |
| Korean | 5 |  | Korean | 7.7 |
| Mandarin | 14 |  | Mandarin | 7.0 |
| Maya | 1 |  | Maya | 7.8 |
| Mongolian | 1 |  | Mongolian | 7.3 |
| Norwegian | 4 |  | Norwegian | 7.2 |
| Persian | 3 |  | Persian | 8.4 |
| Portuguese | 5 |  | Portuguese | 7.8 |
| Romanian | 1 |  | Romanian | 7.9 |
| Russian | 1 |  | Russian | 6.5 |
| Spanish | 23 |  | Spanish | 7.1 |
| Thai | 3 |  | Thai | 6.6 |
| Vietnamese | 1 |  | Vietnamese | 7.4 |
| Zulu | 1 |  | Zulu | 7.3 |
| (blank) |  |  |  |  |
| **Grand Total** | **3722** |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | **StdDevp of imdb\_score** |  | **Row Labels** | **Varp of imdb\_score** |
| Aboriginal | 0.55 |  | Aboriginal | 0.3 |
| Arabic | 0 |  | Arabic | 0.0 |
| Aramaic | 0 |  | Aramaic | 0.0 |
| Bosnian | 0 |  | Bosnian | 0.0 |
| Cantonese | 0.324509048 |  | Cantonese | 0.1 |
| Czech | 0 |  | Czech | 0.0 |
| Danish | 0.43204938 |  | Danish | 0.2 |
| Dari | 0.1 |  | Dari | 0.0 |
| Dutch | 0.329983165 |  | Dutch | 0.1 |
| English | 1.051741952 |  | English | 1.1 |
| Filipino | 0 |  | Filipino | 0.0 |
| French | 0.514240566 |  | French | 0.3 |
| German | 0.67535176 |  | German | 0.5 |
| Hebrew | 0 |  | Hebrew | 0.0 |
| Hindi | 0.716658915 |  | Hindi | 0.5 |
| Hungarian | 0 |  | Hungarian | 0.0 |
| Indonesian | 0.3 |  | Indonesian | 0.1 |
| Italian | 1.200666482 |  | Italian | 1.4 |
| Japanese | 0.920279077 |  | Japanese | 0.8 |
| Kazakh | 0 |  | Kazakh | 0.0 |
| Korean | 0.509901951 |  | Korean | 0.3 |
| Mandarin | 0.737930089 |  | Mandarin | 0.5 |
| Maya | 0 |  | Maya | 0.0 |
| Mongolian | 0 |  | Mongolian | 0.0 |
| Norwegian | 0.497493719 |  | Norwegian | 0.2 |
| Persian | 0 |  | Persian | 0.0 |
| Portuguese | 0.875442745 |  | Portuguese | 0.8 |
| Romanian | 0 |  | Romanian | 0.0 |
| Russian | 0 |  | Russian | 0.0 |
| Spanish | 0.879883521 |  | Spanish | 0.8 |
| Thai | 0.368178701 |  | Thai | 0.1 |
| Vietnamese | 0 |  | Vietnamese | 0.0 |
| Zulu | 0 |  | Zulu | 0.0 |
| **Grand Total** | **1.056748617** |  | **Grand Total** | **1.1** |

Here we can see that English has the highest number of movies.

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

|  |  |
| --- | --- |
| Row Labels | Average of imdb\_score |
| Akira Kurosawa | 8.7 |
| Alfred Hitchcock | 8.5 |
| Asghar Farhadi | 8.4 |
| Charles Chaplin | 8.6 |
| Christopher Nolan | 8.4 |
| Damien Chazelle | 8.5 |
| Majid Majidi | 8.5 |
| Richard Marquand | 8.4 |
| Sergio Leone | 8.4 |
| Tony Kaye | 8.6 |

|  |  |
| --- | --- |
| **Large** | 8.7 |
| **Percentage Rank** | 0.994 |
| **Percentile Rank** | 0.994 |

So the directors having 8.7 IMDB rating has higher impact of getting success on box office.

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

|  |  |
| --- | --- |
|  |  |
| Highest Profit | Name |
| 523505847 | Avatar |
|  |  |
|  |  |
| Correlation |  |
| 0.09828957 |  |

So the highest profit movie is Avatar with the profit of 523505847

To check the correlation between budget and earning we need to find the correlation between then.

From the above table we can see that the correlation between them is close to 0 which means there is no direct relationship between them, even a low budget film earns a good amount of money.

Result: So from the IMDB analysis we can conclude that a good movie can be made by any director and any budget, it depends on the story and the audience to get it succeeded and we have English movies more as compared to other languages.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*THE END\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*