IMDB Movie Analysis

Project Description:

The Internet Movie Database (IMDB) is the source of the movie data which is thoroughly examined in the IMDB Movies Data Analysis project. Finding interesting patterns and trends in several areas of the film industry, such as genres, lengths, languages, directors, and budgets, is the main goal. The study methodically investigates each category to offer a thorough grasp of the cinematic world, using Microsoft Excel for data processing and analysis. The results are provided in an organized report which was created with Microsoft Word and carefully recorded, providing insightful information for scholars, filmmakers, and moviegoers.

Approach:

1.Gathering and Preparing Data

Data Importation: Launched Microsoft Excel and loaded the IMDB dataset for analysis.

Data cleaning: Used the proper approximation techniques and eliminated unnecessary entries to handle partial or

missing data. To preserve the accuracy of data, duplicate records were removed.

Data Validation: Checked that the information entered is accurate. A preliminary analysis was done to find and fix any anomalies or outliers.

Reporting: In the compilation of the report, the IMDB Movies Data Analysis findings are meticulously organized into a structured document using Microsoft Word. Each section of the report is dedicated to a specific area of analysis, with clear and concise narratives explaining the insights derived from the data. Visual aids, including charts and graphs generated in Excel, are seamlessly integrated to support and enhance the written analysis

Tech Stack Used:

For all data-related operations, such as data cleansing, analysis, and visualization, Microsoft Excel is used in the project. Effective visualization of data and insight extraction are achieved through the utilization of Excel's robust functions, pivot tables, and charting features. The final report is created using Microsoft Word and includes integrated charts and graphs to graphically support the findings. When Word and Excel are used together, there is a more efficient process for processing, evaluating, and presenting the data in a way that is both readable and professional.

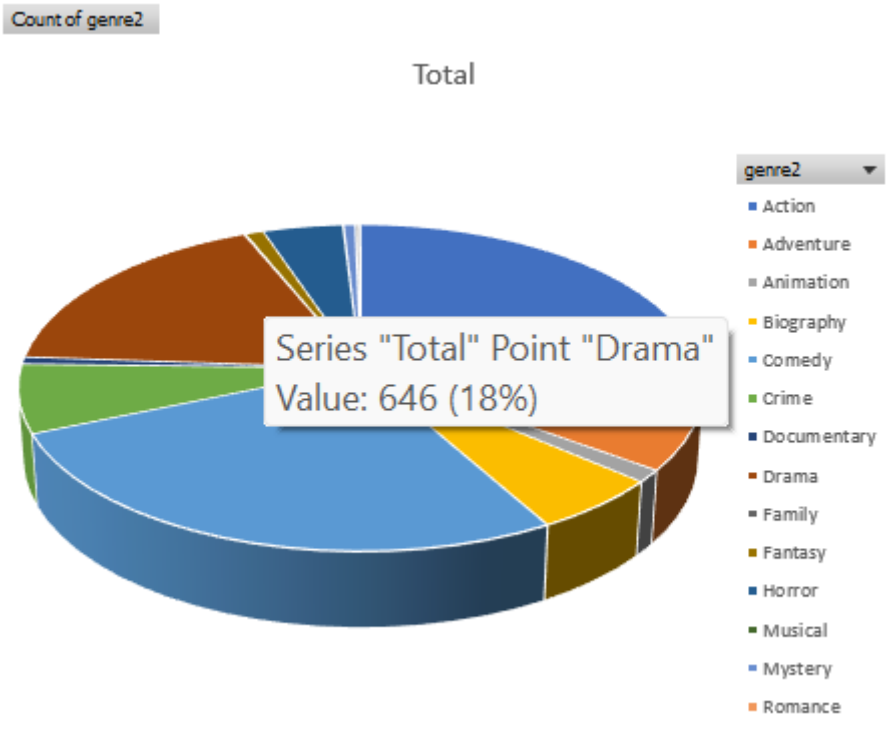
Insights:

1.Movie Genre Analysis:

By analyzing the IMDB movie genre dataset, it was found that action, adventure, animation, biography, etc are top-rated genres in the IMDB database.

Row Labels	Count of genre2
Action	924
Adventure	358
Animation	45
Biography	203
Comedy	980
Crime	248
Documentary	26
Drama	646
Family	3
Fantasy	35
Horror	155
Musical	2
Mystery	22
Romance	1
Sci-Fi	7
Thriller	1
Western	2
Grand Total	**3658**

Genre Statistical Analysis	
Mean	215.1764706
Median	35
Min	1
Max	980
Variance	106484.7794
Standard Deviation	316.5769557

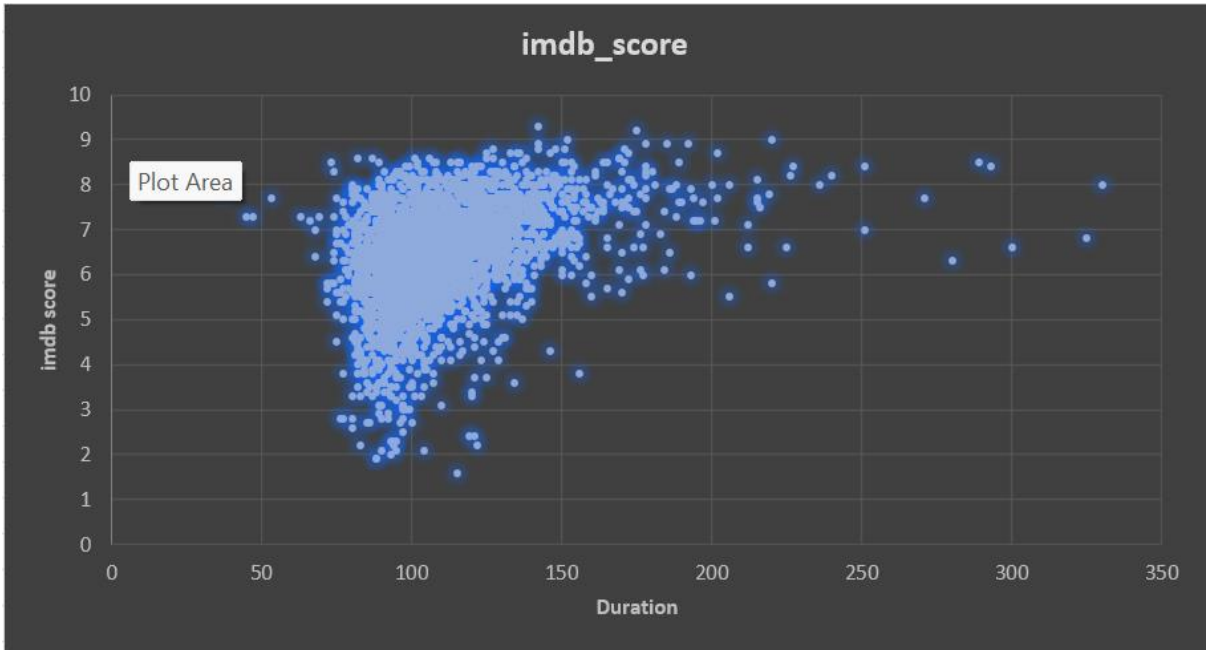


A thorough statistical analysis was done on IMDB genre data which consists of mean, median, max, min, variance, and standard deviation.

2.Movie Duration Analysis:

It was observed that there was a weak correlation between movie duration and IMDB score, showing that movie length alone does not significantly influence audience ratings.

Using a scatter plot it was observed that movies with a duration ranging between 80 mins to 150 mins usually have a high IMDB rating.



A statistical analysis was also done that included mean, median, and standard deviation to understand the duration and IMDB dataset thoroughly.

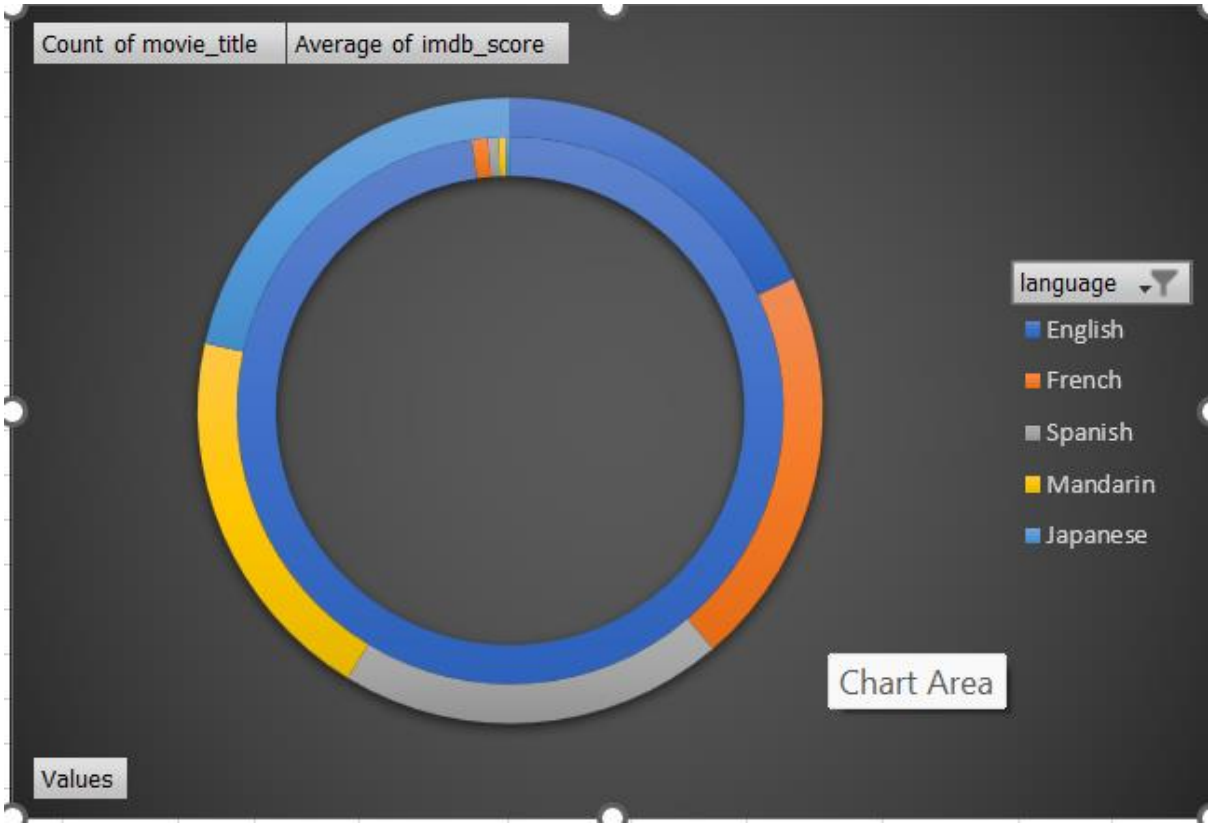
Stastistical Analys	Duration	Imdb scc
Mean	109.9502	6.451278
Median	106	6.6
Standard Deviation	22.68239	1.074283

3.Language Analysis:

After going through the language data of movies it was concluded that most of the movies (approx. 3500) were in English language followed by French and Spanish.

It was found out that the average IMDB score for all the English movies was approx. 6.5, whereas French and Spanish constitute of approx 7 each.

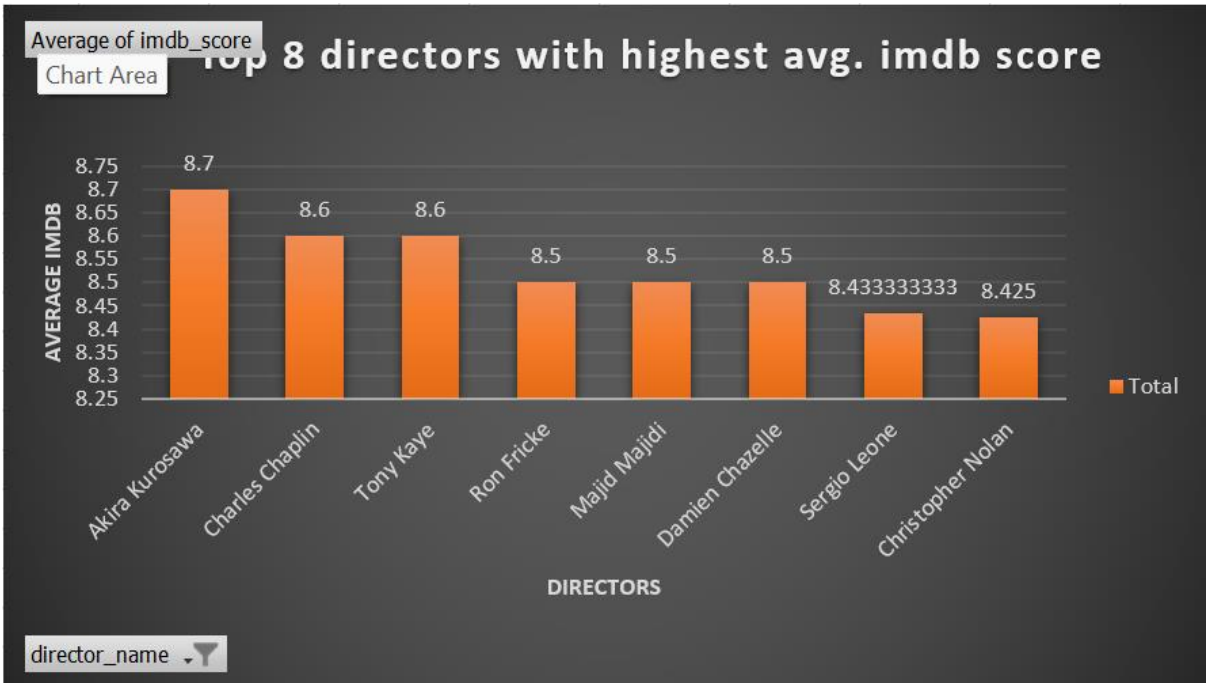
Row Labels	Count of movie_title	Average of imdb_score
English	3501	6.424878606
French	34	7.355882353
Spanish	23	7.082608696
Mandarin	14	7.021428571
Japanese	10	7.66
Grand Total	**3582**	**6.443718593**



4.Director Analysis:

The data was analyzed and found top directors with the highest average IMDB scores. Directors such as Akira Kurosawa, Charles Chaples, Tony Kaye, etc consistently produced movies with high IMDB ratings.

Row Labels	Average of imdb_score
Akira Kurosawa	8.7
Charles Chaplin	8.6
Tony Kaye	8.6
Ron Fricke	8.5
Majid Majidi	8.5
Damien Chazelle	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
Grand Total	**8.476470588**



5.Budget Analysis:

The budget and gross profit data were analyzed to find the relationship between budget and net profit to find a correlation. The financial success also depends on other factors such as Directors, Duration, Actors, Genre, etc.

It was also found that the movies with budget ranging between 2000000000 and 3500000000 have earned more profit than others.

Conclusion:

To sum up, the IMDB Movies Data Analysis project provides insightful information on a range of film industry topics, including trends in the genre, runtimes, language diversity, director impact, and financial concerns. The project successfully finds patterns and connections in the dataset by utilizing Microsoft Word's documentation tools and Microsoft

Excel's powerful analytical capabilities. The results improve our knowledge of the cinematic world and offer helpful data to those involved in the film business. The project makes a substantial addition to the study and analysis of movies since it shows how data analysis may be used to identify trends and guide decision-making processes.

BY:- SOURABH GUPTA

Link of the excel solution sheets:- Link