IMDB MOVIE ANALYTICS

Davidraju Lakkamthoti

AGENDA

Project description

Approach

Tech-stack used

Insights

[Excel file hyperlink](#)

PROJECT DESCRIPTION

The dataset provided is related to IMDB Movies. A potential problem to investigate could be: "What factors influence the success of a movie on IMDB?" Here, success can be defined by high IMDB ratings. The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.

So Data cleaning , Explorative Data analysis with the help of Descriptive Statistics is done to understand the given problem statements.

Excel file hyperlink

APPROACH & TECH STACK USED

Data cleaning is the first step , in which Missing values , blanks , outlier detection is done. Filter function is used to avoid missing values and outliers were detected and omitted ; Pivot tables are majorly used for the analysis.
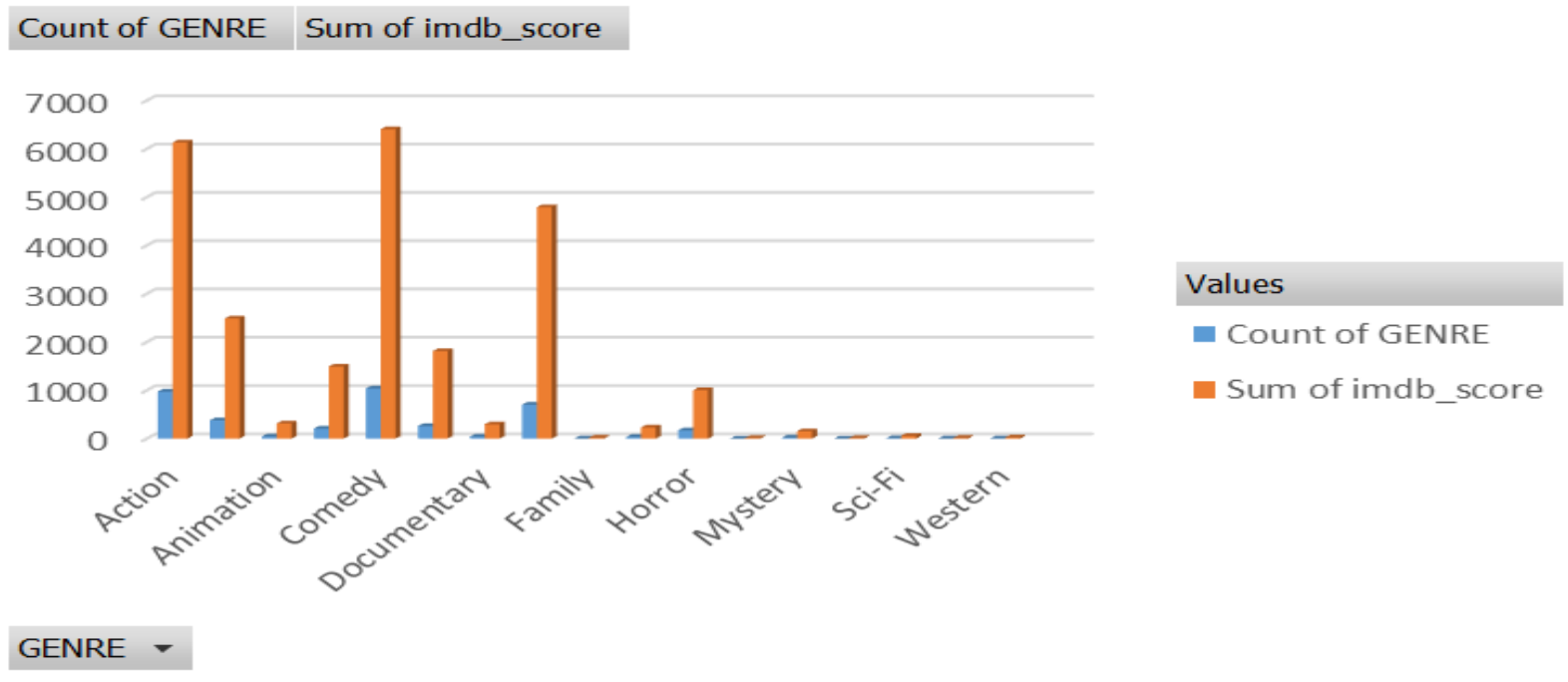
All the irrelevant data is omitted and cleaned data was used for study.

Then Tasks were answered using excel functions.

Microsoft Excel is used for the whole project.

INSIGHTS

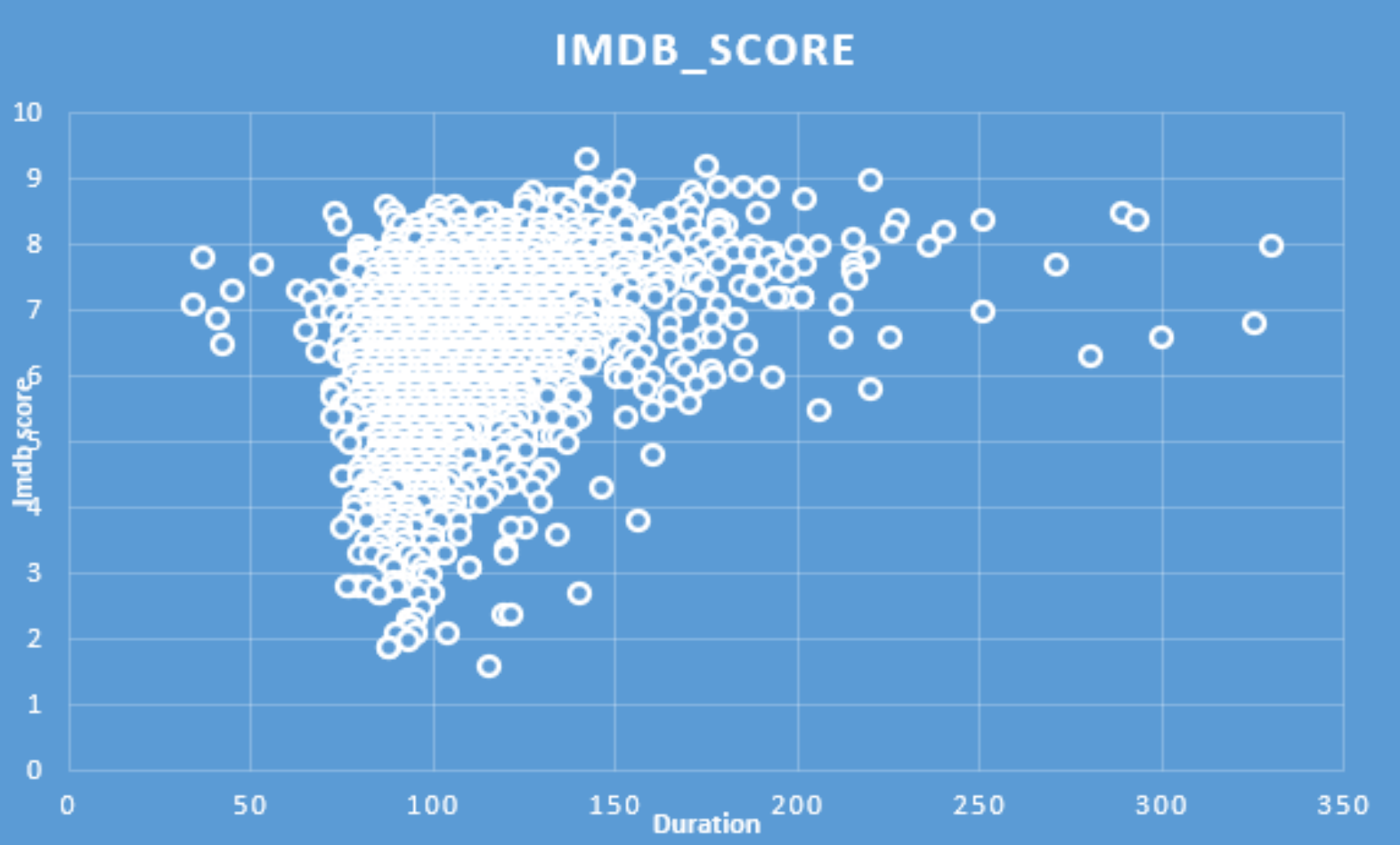**Movie Genre Analysis:** Analyze the distribution of movie genres and their impact on the IMDB score.

Descriptive Analysis of Genre	
Mean	230.5294118
Standard	84.04262389
Median	43
Mode	3
Standard	346.5166153
Sample V.	120073.7647
Kurtosis	1.475177384
Skewness	1.629576652
Range	1039
Minimum	2
Maximum	1041
Sum	3919

We could see ACTION & COMEDY are the most popular Genres . Their count is higher, so the IMDB rating also is high.

It means People are mostly entertained by Action, Comedy And Drama Genres . These three genres also seems to be most successful as they were liked by the people.

Descriptive analysis of genre count column: Mean value is not significant , as the distribution highly skewed with the value 1.6.   Excel file hyperlink

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

IMDB_SCORE

Scatterplot for Imdb Score vs Duration of movie is shown above.

Mean duration is 109.78 minutes.

Median is 106 minutes.

Scatterplot clearly shows, movies with 100 to 150 minutes duration has got highest imdb score. It is also evident that most of the movies were made with a duration of 109 minutes. It is ideal to make a movie with such duration to stand successful in movie making.

Excel file hyperlink

Language	Mean	STD DEV
Aboriginal	6.95	0.55
Arabic	7.20	0.00
Aramaic	7.10	0.00
Bosnian	4.30	0.00
Cantonese	7.24	0.41
Czech	7.40	0.00
Danish	7.90	0.43
Dari	7.50	0.10
Dutch	7.57	0.33
Dzongkha	7.50	0.00
English	6.41	1.06
Filipino	6.70	0.00
French	7.29	0.55
German	7.69	0.62
Hebrew	7.50	0.36
Hindi	6.76	1.05
Hungarian	7.10	0.00
Icelandic	6.90	0.00
Indonesian	7.90	0.30
Italian	7.19	1.07

INSIGHTS

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

Most common language used in movies is ENGLISH, Total count was 3735. Next language is French with a total count of 37.

In this analysis we can observe each language mean and their standard deviation, If the std dev is close to 1 means 68% of data points will fall under one std deviation of the mean. English, Hindi and Italian shows such behavior.

AS English language count is highest , based on the results , we can conclude in English language case.

Excel file hyperlink

Top 10 Directors	Average of imdb_score
Charles Chaplin	8.60
Tony Kaye	8.60
Alfred Hitchcock	8.50
Damien Chazelle	8.50
Majid Majidi	8.50
Ron Fricke	8.50
Sergio Leone	8.43
Christopher Nolan	8.43
Asghar Farhadi	8.40
Marius A. Markevicius	8.40
PERCENTILE	8.40

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

Percentile value tells you the relative position of a data point in the whole dataset.

Here in the given case , 8.4 is the percentile value, it means any director if he scores 8.4 imdb score for his movie , He would be one of the top director. So percentile value can define whether the given director score is in top 10 list or not. Charles chaplin and tony kaye are top two directors.

gross	budget	Profit	movie_title	CORRELATION
28341469	25000000	3341469	The Shawshank Redemption	0.1021773361
134821952	6000000	128821952	The Godfather	
533316061	185000000	348316061	The Dark Knight	
57300000	13000000	44300000	The Godfather: Part II	Avatar movie profit
377019252	94000000	283019252	The Lord of the Rings: The Return of the King	523505847
96067179	22000000	74067179	Schindler's List	
107930000	8000000	99930000	Pulp Fiction	
6100000	1200000	4900000	The Good, the Bad and the Ugly	
292568851	160000000	132568851	Inception	
313837577	93000000	220837577	The Lord of the Rings: The Fellowship of the Ring	
37023395	63000000	-25976605	Fight Club	
329691196	55000000	274691196	Forrest Gump	
290158751	18000000	272158751	Star Wars: Episode V - The Empire Strikes Back	

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

[Excel file hyperlink](#)

The correlation value for Movie budget and gross earning is 0.1. It means there is no direct relationship between budget and earnings.

Maximum highest profit margin is earned by avatar movie. Which is observed using max function and xmatch function.

THANK YOU

Davidraju Lakkamthoti

ge19lakkamthoti@mse.ac.in