

IMDB Movie Analysis

Presented By: Harsh Dikshit



Project Description

In this project, we are provided with an IMDB movie dataset. We have to investigate the dataset and draw meaningful insights from it like, "What factors influence the success of a movie in IMDB?". This may help directors, investors, and producers to understand what makes a movie a great success and apply those insights in their future projects.

Approach

My approach towards this project was to, firstly go through the dataset and provided and understand all the values in it. After that, I will clean the dataset by removing the blanks and duplicates value. This will make my dataset more accurate. Now, i will go through the questions asked by the team and provide answers to their question by generating meaningful insights from the dataset.



Tech Stack Used

I have used MS Excel 2019. It provides us with handy tools and mathematical functions, which help us solve mathematical problems. This tool also provides us with many charts to display our insights more effectively.



Insights

Through this project, I understand how IMDB movie data helps the directors, producers, and investors understand the factors influencing the success of the movie. This project helped me in getting advanced knowledge of MS Excel. After working on this project, I made an observation about how movie directors/producers use this data to make their future projects more successful.



Result

Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics of the IMDB scores.

Geners	Count	Average	Max	Min	Standard Deviation	Variance
Drama	1876	6.79	9.3	2.1	0.89	0.79
Comedy	1455	6.19	8.8	1.9	1.03	1.07
Thriller	1105	6.38	9	2.7	0.97	0.94
Action	951	6.29	9	2.1	1.03	1.07
Romance	851	6.44	8.5	2.1	0.96	0.91
Adventure	773	6.45	8.9	2.3	1.11	1.24
Crime	704	6.54	9.3	2.4	0.98	0.96
Fantasy	504	6.29	8.9	2.2	1.13	1.27
Sci-Fi	492	6.32	8.8	1.9	1.16	1.34
Family	440	6.22	8.6	1.9	1.16	1.35

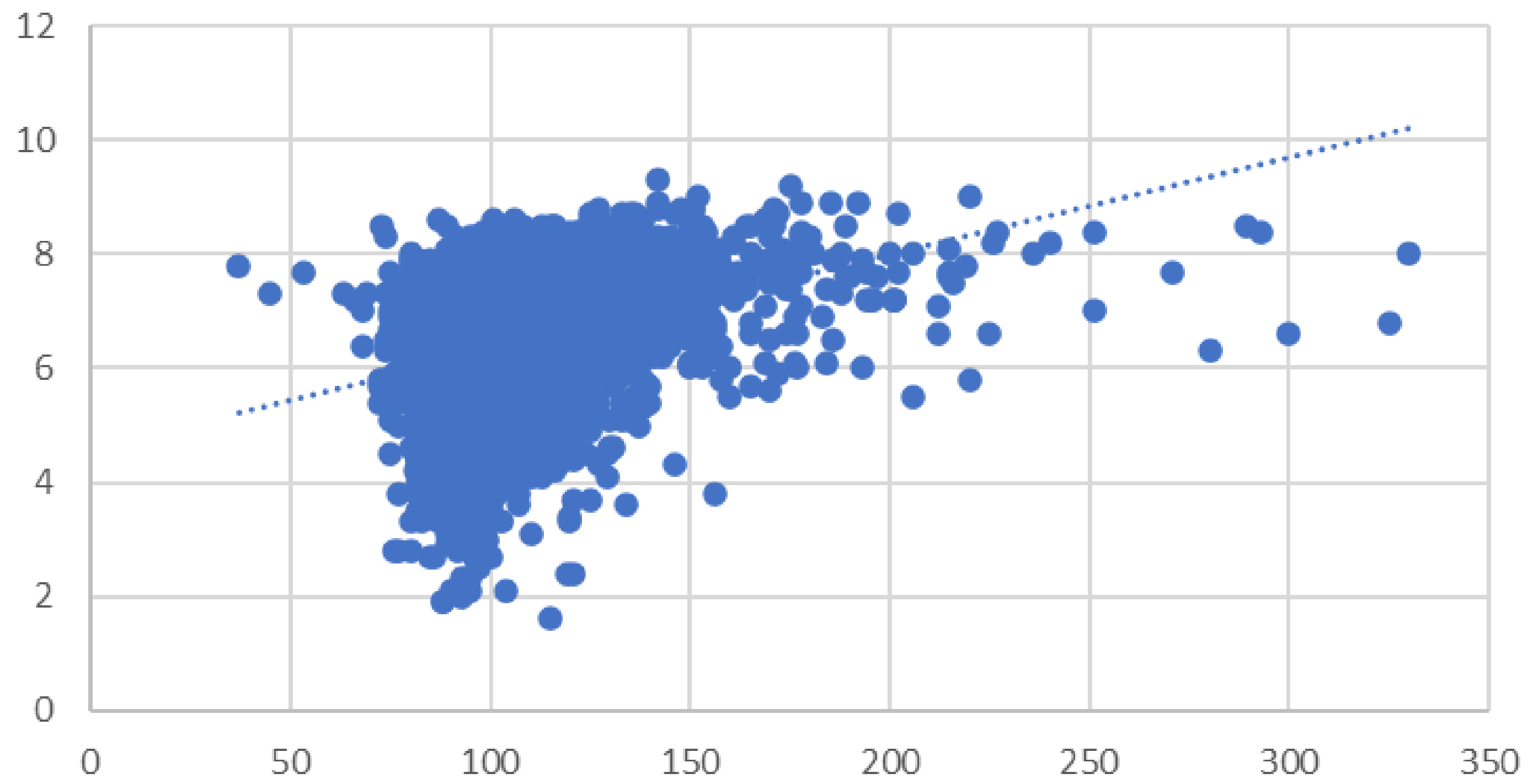
Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

110.2635
Average

106
Median

22.67832498
Standard Deviation

imdb_score



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

Languages	Count of movies	Avg of IMDB score	Std Dev of IMDB score
Cantonese	7	7.342857143	0.350509833
English	3566	6.427509815	1.047903308
French	34	7.355882353	0.519435111
German	10	7.77	0.711883261
Hindi	5	7.22	0.801249025
Italian	7	7.185714286	1.155318962
Japanese	10	7.66	0.990173947
Korean	5	7.7	0.570087713
Mandarin	14	7.021428571	0.765786244
Portuguese	5	7.76	0.978774744
Spanish	23	7.082608696	0.860577065

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

Director Name	Avg IMDB Score
Akira Kurosawa	8.7
Charles Chaplin	8.6
Tony Kaye	8.6
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Ron Fricke	8.5
Sergio Leone	8.4333333333
Christopher Nolan	8.425
Asghar Farhadi	8.4
Richard Marquand	8.4

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

Movie with highest profit: Avatar (523505847)

Correlation coefficient: 0.098318102



File Link

<https://docs.google.com/spreadsheets/d/1YQThUjZIZgmlgyR54KowU-gDQKjjq6PC/edit?usp=sharing&ouid=117831886349301749092&rtpof=true&sd=true>

Thank You!