Trainity Project Report

IMDB Movie Analysis

Approach:

Before beginning the analysis, I carried out data cleaning and preparation steps to ensure accuracy and relevance of the findings. Here's a brief overview of the approach followed:

- **Data Cleaning:** Removed duplicate rows and unnecessary columns. - Filled missing values where appropriate or excluded incomplete records. Standardized column names for easier reference. - Formatted data types correctly (e.g., numeric values for budgets and scores).

- **Data Categorization:** Separated multiple genres into primary categories for clarity. Grouped languages and directors for comparative analysis.

- **Analysis Setup:** Used pivot tables and Excel functions to summarize data. - Applied conditional formatting and trend lines to visualize patterns. - Used formulas like AVERAGE, MEDIAN, MODE, STDEV, CORREL, and others.

Once the data was clean and structured, I moved on to the main analysis tasks, which are described in the following sections.

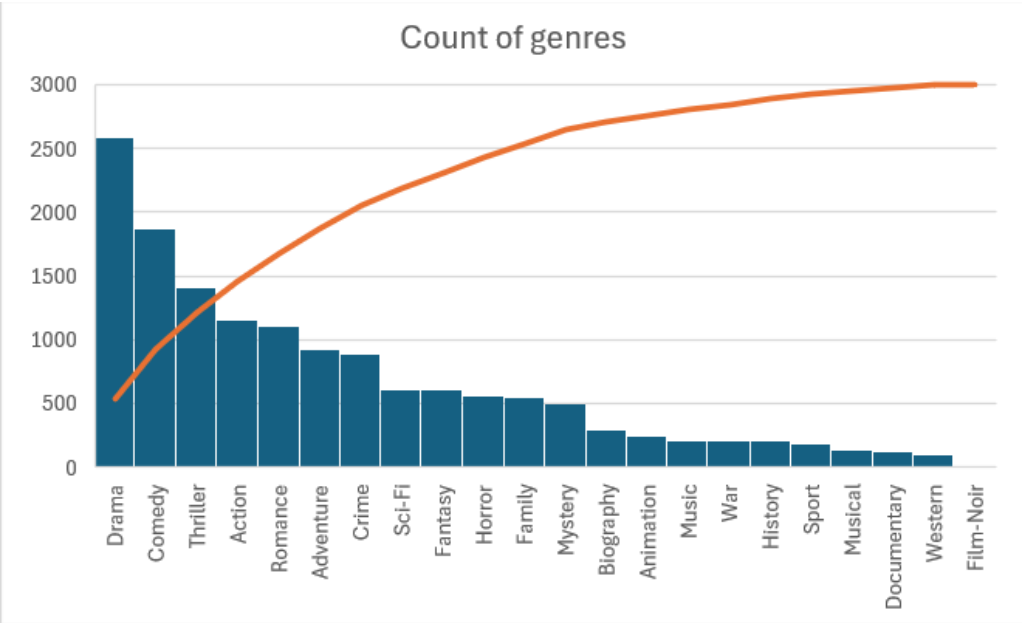
Tools and Techniques Used:

- Microsoft Excel (Formulas, Pivot Tables, Charts)
- Google Drive (Data Sharing)
- Loom (For Presentation Recording and Sharing)
- Descriptive Statistics (Mean, Median, Mode, Variance, Standard Deviation)

Data Analytics Tasks:

A. Movie Genre Analysis:

Task: 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.



Explanation:
First, I separated genres listed together in a single cell into individual entries. Then, using the COUNTIF function in Excel, I calculated the frequency of each genre. Among all, Drama had the highest count, indicating it was the most represented genre in the dataset. Drama had a wide range of scores but generally scored higher.

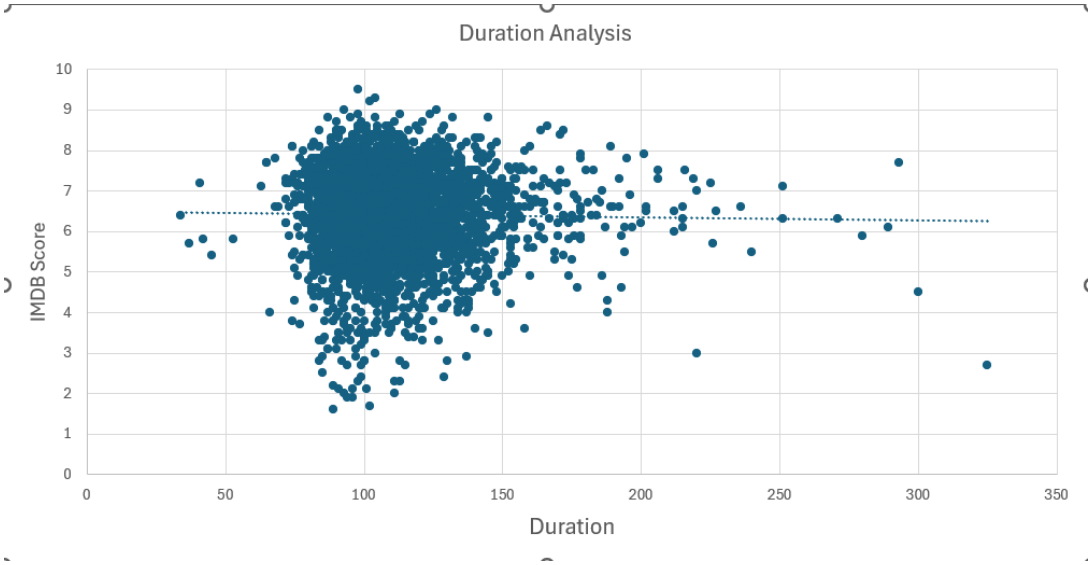
New_genre	Count of genre
Action	1153
Adventure	923
Drama	2594
Music	214
Horror	565
Romance	1107
Comedy	1872
Crime	889
Biography	293
Fantasy	610
Mystery	500
Musical	132
Thriller	1411
History	207
Animation	242
Family	546
Western	97
Sci-Fi	616
Sport	182
War	213
Documentary	121
Film-Noir	6
Grand Total	14493

AVERAGE	658.77
MEAN	659
MODE	0
MEDIAN	523
ST.DEV	649.0
MAX	2594
MIN	6
VAR	421220

Insight:
Drama and Biography genres tend to receive better ratings compared to genres like Horror or Thriller. With the count of 2594. And the least genre is Film-Noir with the count of 6. As compared to the other.

B. Movie Duration Analysis:

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



Explanation:
Most movies ranged between 80-130 minutes. A scatter plot and trend line showed a weak positive correlation between longer durations and higher ratings.

AVERAGE	109
MEDIAN	105
ST.DEV	22.50

Insight:
Movies with very short or excessively long durations typically received mixed reviews.

C. Language Analysis:

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

English	3605
Filipino	1
French	36
German	13
Hebrew	3
Hindi	9
Hungarian	1
Icelandic	1
Indonesian	2
Italian	8
Japanese	13
Kazakh	1
Korean	5
Mandarin	14
Maya	1
Mongolian	1
None	1
Norwegian	4
Persian	5
Portuguese	5
Romanian	1
Russian	1
Spanish	26

Mean	Median	St.d
8	6.5	1.1
6.4		
7.2		
7		
6.2		
6.2		
6.27		
5.6		
6.83		
6.1		
6.42		
6.8		
6.34		
6.38		
5.23		
6.22		
5.1		
6.1		
7.45		
6.3		
6.5		
5.8		
5.52		
6.56		
7		
5.8		
8		
7.03		
5.66		

Explanation:
English dominated the dataset, followed by French, Spanish, and German. English-language movies generally had higher ratings with the **Count of Language** is 3605 and **Mean** of 6.24. I have not mentioned other Median and standard deviations because of repetition of same Median and St.d.

Insight:
English movies are more widely reviewed, possibly explaining their slightly higher average ratings.

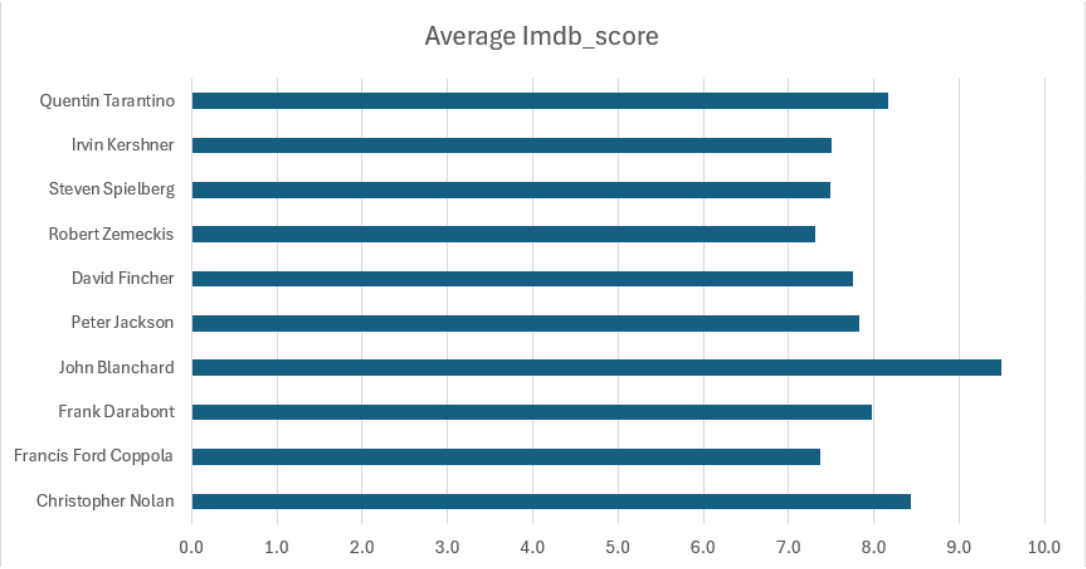
D. Director Analysis:

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

Top 10 Director name	Average Imdb_score
Christopher Nolan	8.4
Francis Ford Coppola	7.4
Frank Darabont	8.0
John Blanchard	9.5
Peter Jackson	7.8
David Fincher	7.8
Robert Zemeckis	7.3
Steven Spielberg	7.5
Irvin Kershner	7.5
Quentin Tarantino	8.2

Large	9.5
Percent rank	1
Percentile	9.5

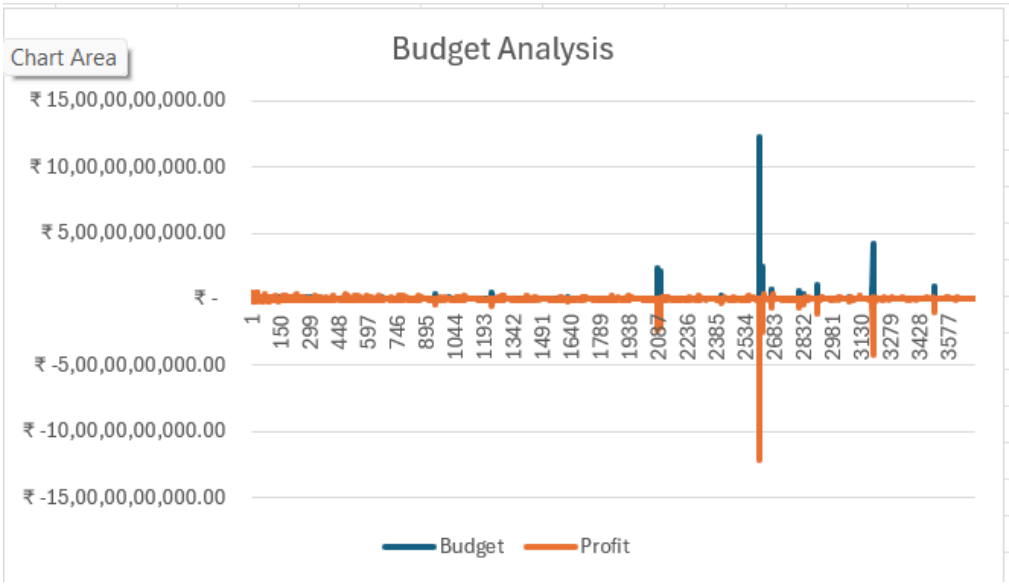
Explanation:
Directors such as John Blanchard and Christopher Nolan received the higher scores. As compared to the other Top Directors. And Using the Percentile We can See that which director's movies having the success rate.



Insight:
Experienced and well-known directors significantly contribute to a movie's success.

E. Budget Analysis:

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



CORRELATION
0.095753134

Highest Profit
523505847

Explanation:

High-budget movies tended to generate more revenue, but there were exceptions. A scatter chart showed moderate correlation.

Insight:

Investing more can result in better returns, but genre, timing, and director also play crucial roles.