IMDB Movie Analysis Report

Submitted by Shahanas Beegam

Tasks

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

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.
Hint: Use Excel's COUNTIF function to count the number of movies for each genre. You might need to manipulate the 'genres' column to separate multiple genres for a single movie. Use Excel's functions like AVERAGE, MEDIAN, MODE, MAX, MIN, VAR, and STDEV to calculate descriptive statistics. Compare the statistics to understand the impact of genre on movie ratings.

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

Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.
Hint: Calculate descriptive statistics such as mean, median, and standard deviation for movie durations. Use Excel's functions like AVERAGE, MEDIAN, and STDEV. Create a scatter plot to visualize the relationship between movie duration and IMDB score. Add a trendline to assess the direction and strength of the relationship.

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

Task: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.
Hint: Use Excel's COUNTIF function to count the number of movies for each language. Calculate the mean, median, and standard deviation of the IMDB scores for each language. Compare the statistics to understand the impact of language on movie ratings.

D. Director Analysis: Influence of directors on movie ratings.

Task: Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.
Hint: Calculate the average IMDB score for each director. Use Excel's PERCENTILE function to identify the directors with the highest scores. Compare the scores of these directors to the overall distribution of scores.

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

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

Hint: Calculate the correlation coefficient between movie budgets and gross earnings using Excel's CORREL function. Calculate the profit margin (gross earnings - budget) for each movie and identify the movies with the highest profit margin using Excel's MAX function.

A. Movie Genre Analysis

1. Genre Distribution Analysis

Data provides insights into the counts of movies across various genres and their statistical measures. This part focused on understanding how the distribution of genres might impact the overall success of movies.

Summary Statistics for Movie Genres:

- **Average Count**: 394.33
- **Median Count**: 214
- **Mode**: 6
- **Maximum Count**: 1623 (Drama)
- **Minimum Count**: 3 (Reality-TV)
- **Variance**: 197,356.91
- **Standard Deviation**: 444.25

Genere	Count
Adventure	471
Animation	182
Biography	42
Comedy	544
Crime	541
Documentary	38
Drama	1623
Family	536
Fantasy	555
Film-Noir	6
History	207
Horror	334
Music	214
Musical	128
Mystery	467
News	4
Reality-TV	3
Romance	1099
Sci-Fi	597
Short	6
Sport	183
Thriller	1385
War	214
Western	85

B. Movie Duration Analysis

Objective: Analyse the distribution of movie durations and its impact on IMDB scores.

Duration (minutes)	IMDB Score
178	7.9
169	7.1
148	6.8
164	8.5

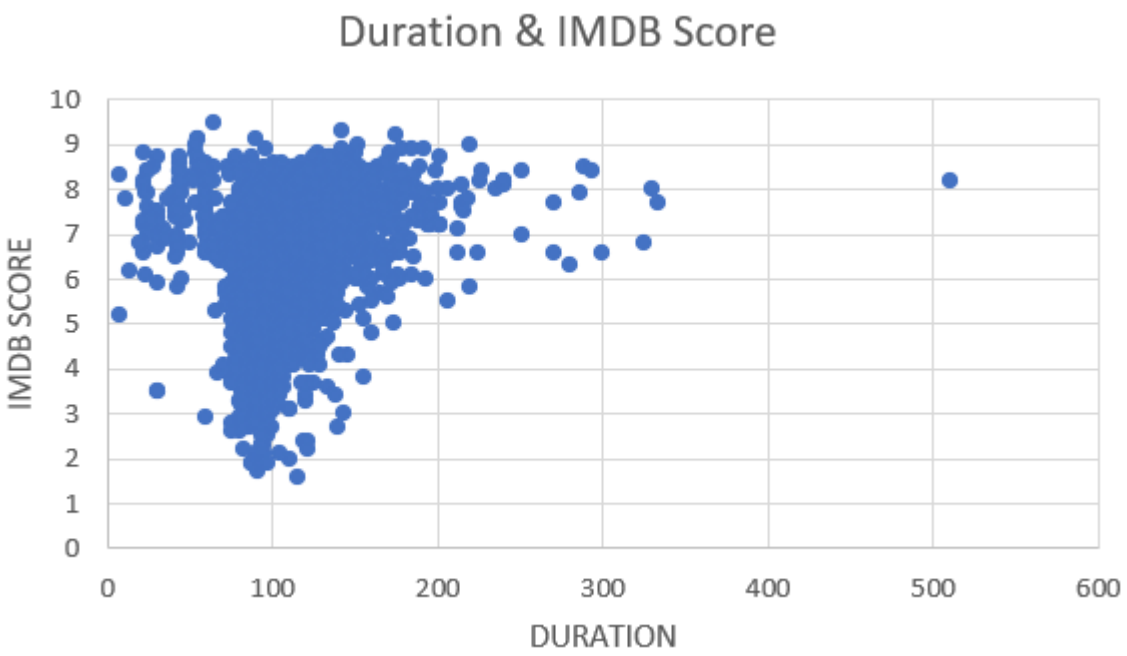
132	6.6
156	6.2
100	7.8
141	7.5
153	7.5
183	6.9
169	6.1
106	6.7
151	7.3
150	6.5
143	7.2
150	6.6
173	8.1
136	6.7

Statistics for Movie Duration:

- **Average Duration**: 107.20 minutes
- **Median Duration**: 103 minutes
- **Standard Deviation**: 25.20
- **Minimum Duration**: 7 minutes (likely an error or placeholder, usually should be ignored)
- **Maximum Duration**: 183 minutes

Insights and Analysis:

1. **Average and Median Durations**:
    - The average duration of movies is approximately 107 minutes.
    - The median duration is 103 minutes, indicating that half the movies are longer than this and half are shorter.
2. **Spread of Durations**:
    - The standard deviation of 25.20 minutes shows moderate variability in movie durations.
3. **Relationship Between Duration and IMDB Score**:



- Scatter plot of movie duration vs. IMDB score with a trendline to show the relationship.

C. Language Analysis

Objective: Examine the distribution of movies based on their language and analyse their impact on IMDB scores using descriptive statistics.

Language	Count
English	4703
Japanese	18
French	73
Mandarin	26
Aboriginal	2
Spanish	40
Filipino	1
Hindi	28
Russian	11
Maya	1

Kazakh	1
Telugu	1
Cantonese	11
Icelandic	2
German	19
Aramaic	1
Italian	11
Dutch	4
Dari	2
Hebrew	5
Chinese	3
Mongolian	1
Swedish	5
Korean	8
Thai	3
Polish	4
Bosnian	1
None	2
Hungarian	1
Portuguese	8
Danish	5
Arabic	5
Norwegian	4
Czech	1
Kannada	1
Zulu	2
Panjabi	1

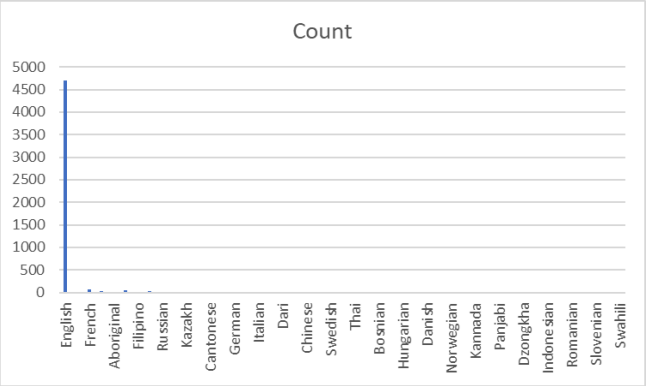
Tamil	1
Dzongkha	1
Vietnamese	1
Indonesian	2
Urdu	1
Romanian	2
Persian	4
Slovenian	1
Greek	1
Swahili	1

Item	Value
Mean	107.0213
Median	2
Standard Deviation	685.0847

Summary Statistics for Languages:

- **Average Count**: 107.02
- **Median Count**: 2
- **Standard Deviation**: 685.08

Insights and Analysis:



1. **Dominant Language**:
    - English is the predominant language with 4703 movies, significantly higher than any other language.

D. Director Analysis

Objective: Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies.

Top Directors

Director	Average IMDB Score
John Blanchard	9.5
Cary Bell	8.7
Mitchell Altieri	8.7
Sadyk Sher-Niyaz	8.7
Charles Chaplin	8.6
Mike Mayhall	8.6
Damien Chazelle	8.5
Majid Majidi	8.5
Raja Menon	8.5
Ron Fricke	8.5

Summary Statistics for Directors' IMDB Scores:

- **Average IMDB Score**: 6.18
- **Median IMDB Score**: 6.3
- **Standard Deviation**: 1.16
- **Percent Rank**: 1
- **Percentile (Top Score)**: 9.5

Insights and Analysis:

1. **Top Directors**:
    - The top directors listed have average IMDB scores ranging from 8.5 to 9.5, indicating their significant impact on movie success.
2. **Comparison with Overall Distribution**:
    - The average IMDB score for these directors (9.5) is well above the general average (6.18), suggesting that these directors consistently produce highly rated movies.

E. Budget Analysis

Objective: Explore the relationship between movie budgets and their financial success.

Data

Gross	Budget	Profit
760505847	237000000	523505847
309404152	300000000	9404152
200074175	245000000	-44925825
448130642	250000000	198130642
73058679	263700000	-190641321
336530303	258000000	78530303
200807262	260000000	-59192738
458991599	250000000	208991599
301956980	250000000	51956980
330249062	250000000	80249062
200069408	209000000	-8930592
168368427	200000000	-31631573
423032628	225000000	198032628
89289910	215000000	-125710090
291021565	225000000	66021565
141614023	225000000	-83385977

Item	Value
Mean	6.176387
Median	6.3
Standard Deviation	1.159526
Percent Rank	1
Percentile	9.5

Summary Statistics for Profit:

- **Correlation between Gross and Budget**: 0.1022 (low positive correlation)
- **Maximum Profit**: $523,505,847

Insights and Analysis:

1. **Profit Analysis**:
    - The correlation of 0.1022 indicates a very weak positive relationship between budget and gross profit, suggesting that higher budgets do not strongly predict higher gross earnings.

Download Excel File