IMDB Movie Analysis

Done by
-Ayithapu Sai Mahathi

PROJECT DESCRIPTION

The project is about finding out valuable insights that can help stakeholders make informed decisions. We analyze this data on the following points:

A. Movie Genre Analysis
B. Movie Duration Analysis
C. Language Analysis
D. Director Analysis
E. Budget Analysis

Software used:

- Microsoft Excel 2307

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

Task A: 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.

- First process involves cleaning the data. So dropping the columns which we have no use for analysis.
- Columns like **color, director_facebook_likes, actor_3_facebook_likes, actor_2_name, actor_1_facebook_likes, cast_total_facebook_likes, actor_3_name, facenumber_in_poster, plot_keywords, movie_imdb_link, content_rating, actor_2_facebook_likes, aspect_ratio, movie_facebook_likes** are irrelevant data. It needs to be dropped.
- Now we need to remove the rows which contains null values. Then we need to remove duplicates from dataset.
- Then we will separate multiple genres and use COUNTIF function to count the number of movies for each genre.
- Then we will use Excel's functions like AVERAGE, MEDIAN, MODE, MAX, MIN, VAR, and STDEV to calculate descriptive statistics.

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

First I have actually used a power pivot table for the columns imdb score,movie title and the genre now in the genre I have done the split data using a delimiter " | " and splited the data and then unpivoted the columns and I got the column names as Drama and all .Then I have created a pivot table then summarized it using count,min,max,standard deviation,variance

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

Output/Result:-

Genres	StdDevp of imdb_score2	Varp of imdb_score2	Max of imdb_score2	Min of imdb_score2	Average of imdb_score2	Count of movie_title
Action	1.037802316	1.077033647	9	2.1	6.285989305	935
Adventure	1.122898367	1.260900742	8.6	2.3	6.561307902	367
Animation	0.961963277	0.925373346	8	4.5	6.763043478	46
Biography	0.697826804	0.486962249	8.9	4.5	7.151941748	206
Comedy	1.036284472	1.073885506	8.8	1.9	6.163609756	1025
Crime	0.867042697	0.751763039	9.3	3.3	6.945238095	252
Documentary	1.444798861	2.08744375	8.5	1.6	6.9175	40
Drama	0.905269207	0.819512338	8.8	2.1	6.821745562	676
Family	0.993310962	0.986666667	7.9	5.7	6.5	3
Fantasy	0.881214375	0.776538776	7.9	4.3	6.234285714	35
Horror	1.004546494	1.009113659	8.5	2.3	5.813461538	156
Musical	0.45	0.2025	7.2	6.3	6.75	2
Mystery	1.088111944	1.183987603	8.5	3.3	6.631818182	22
Romance	0.45	0.2025	7.1	6.2	6.65	2
Sci-Fi	0.964931992	0.93109375	8.2	5	6.5875	8

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

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

- First we will select column **duration** and **imdb_score.**
- Then we will use Excel's functions like AVERAGE, MEDIAN, and STDEV to calculate descriptive statistics.

Formulas:-

Mean: =AVERAGE(A:A)

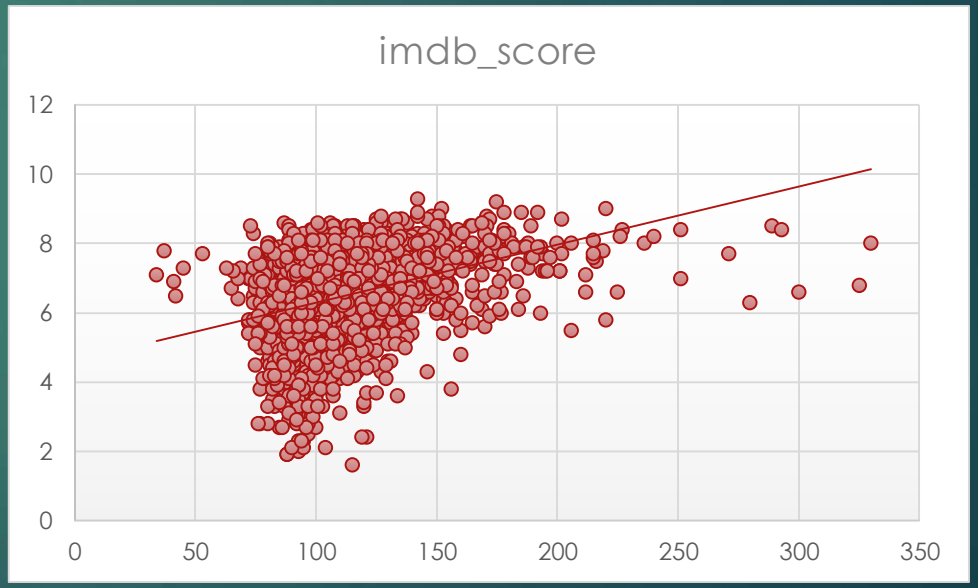
Median: =MEDIAN(A:A)

Standard deviation: =STDEV.S(A:A)

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

Output/Result:-

Mean	standard Deviation	Median
109.9241	22.75069304	106



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

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

- First we will select Column **language** and **imdb_score.**
- Then we will use COUNTIF function to count the number of movies for each language.
- Using AVERAGE, MEDIAN, and STDEV function we will calculate Mean, Median and Standard Deviation of IMDB Scores for each language.

Formulas:

Count: =COUNTIFS('cleaned data'!$J$2:$J$3849, J2)

Mean: =AVERAGE(IF('cleaned data'!$J$2:$J$3849=J2, 'cleaned data'!$N$2:$N$3849))

Median: =MEDIAN(IF('cleaned data'!$J$2:$J$3849=J2, 'cleaned data'!$N$2:$N$3849))

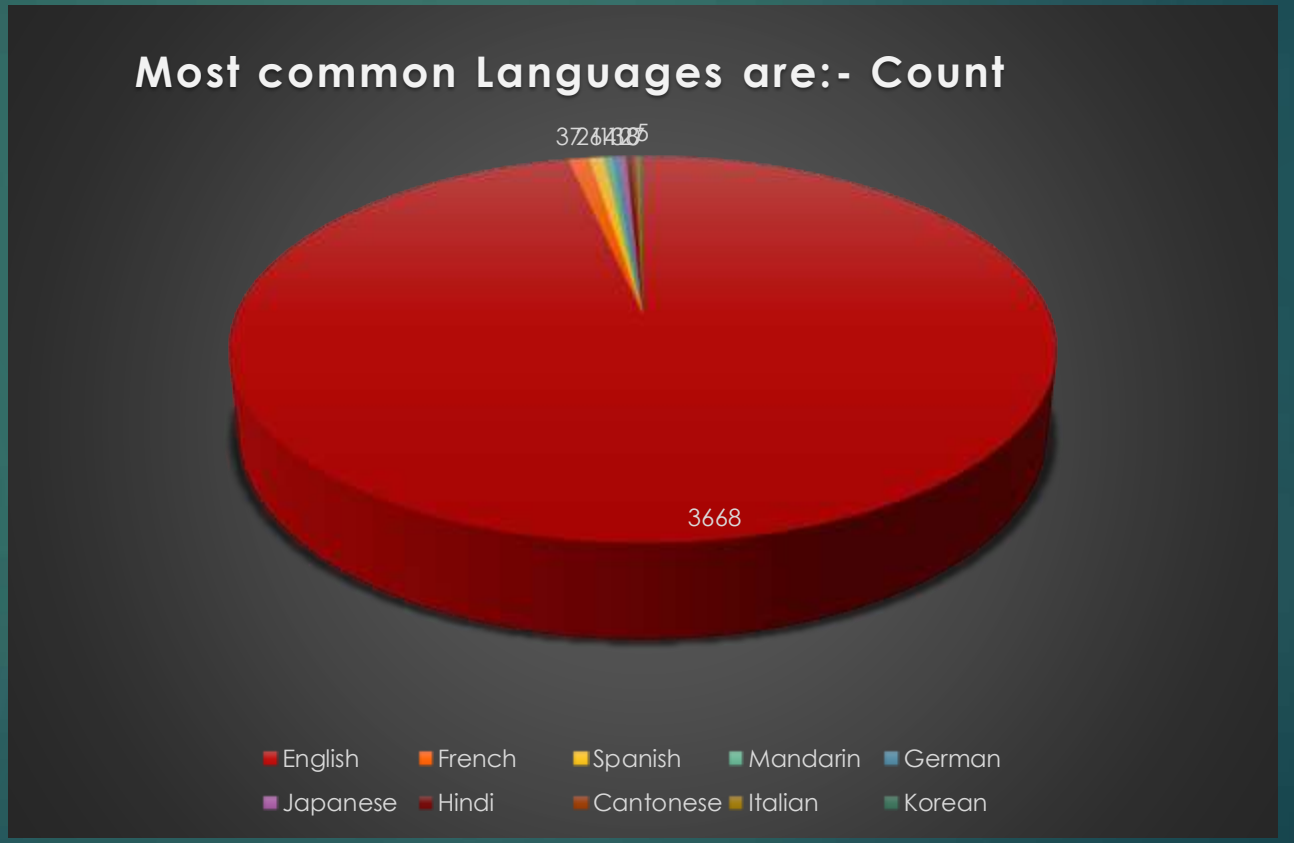
Standard Deviation: =STDEV.S(IF('cleaned data'!$J$2:$J$3849=J2, 'cleaned data'!$N$2:$N$3849))

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

Output/Results:-

Most common Languages are:-				
Language	**Count**	**Mean**	**Median**	**Standard Deviation**
English	3668	6.42391	6.5	1.048750752
French	37	7.28649	7.2	0.561328861
Spanish	26	7.05	7.15	0.826196103
Mandarin	14	7.02143	7.25	0.765786244
German	13	7.69231	7.7	0.640912811
Japanese	12	7.625	7.8	0.899621132
Hindi	10	6.76	7.05	1.111755369
Cantonese	8	7.2375	7.3	0.440575922
Italian	7	7.18571	7	1.155318962
Korean	5	7.7	7.7	0.570087713

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



Director Analysis: Influence of directors on movie ratings.

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

- We will select column **director_name** and **imdb_score.**
- Then we will use AVERAGE function to Calculate the average IMDB score for each director.
- Then we will calculate percentrank and use PERCENTILE function to identify the directors with the highest scores.

Formulas: -

Average: =AVERAGE(IF('cleaned data'!$A$2:$A$3849=A2, 'cleaned data'!$N$2:$N$3849))

Percentile: =PERCENTILE(H2:H11, H15)

Director Analysis: Influence of directors on movie ratings.

Output/Results:-

Director name	Average of imdb_score
Akira Kurosawa	8.1
Alejandro AmenÃ¡bar	7.633333333
Alejandro G. IÃ±Ã¡rritu	7.84
Alex Garland	7.7
Alex Gibney	7.7
Alfonso CuarÃ³n	7.8
Alfred Hitchcock	8.5
Andrew Haigh	7.7
Andrew Stanton	7.733333333
Anna Muylaert	7.9
Ari Folman	8
Asghar Farhadi	8.4
Ben Affleck	7.65
Bernardo Bertolucci	7.8
Billy Wilder	8.3
Brian Henson	7.7
Caroline Link	7.7
Chan-wook Park	7.633333333

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

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

- First we will calculate profit margin for each movie by subtracting budget value from gross value.
- We will use CORREL function to calculate correlation coefficients between movie budgets and gross earnings.

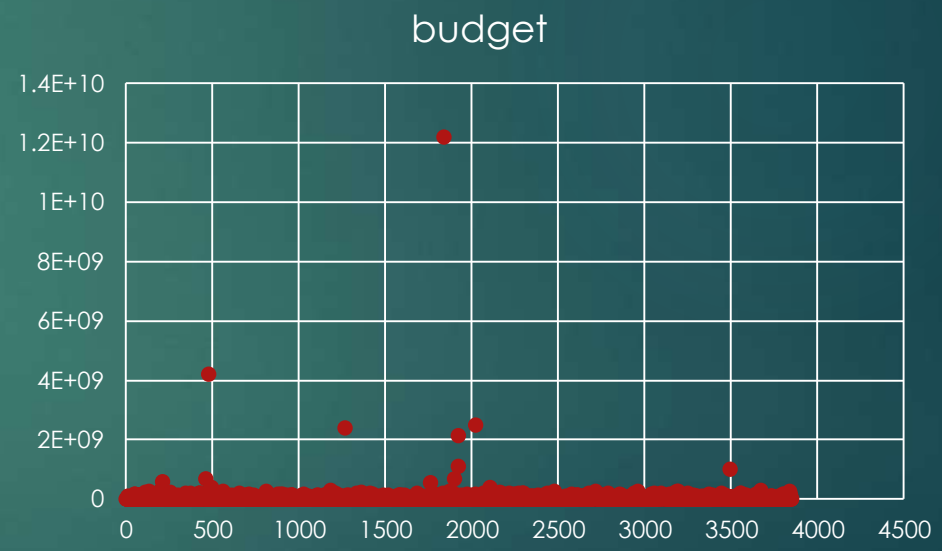
Budget Analysis: Explore the relationship between movie budgets and their financial success.

Output/Results:

Maximum	523505847	AvatarÂ

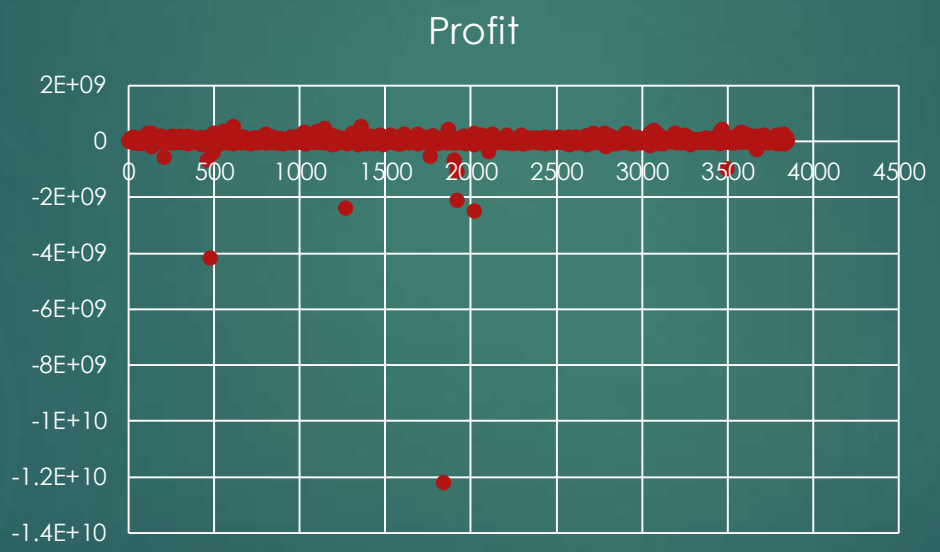
0.100850218	Correlation

Vertical (Value) Axis



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

Output/Results:



Link for my work
C:\Users\91939\OneDrive\Desktop\IMDB_Movies final.csv