IMDB Movie Analysis

Final Project-1

Description:

Problem Statement: 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.

Google Excel sheet Link:https://drive.google.com/drive/folders/1_DFnB6DwOB798PRT1MujQHmqOVNCtusz?usp=sharing

Loom Video Link:https://www.loom.com/share/91dcd6a041af4db59b8a3463832c8e12?sid=459d6893-4e4b-4ab8-afcc-4b42e579f398

Data Analytics Tasks:

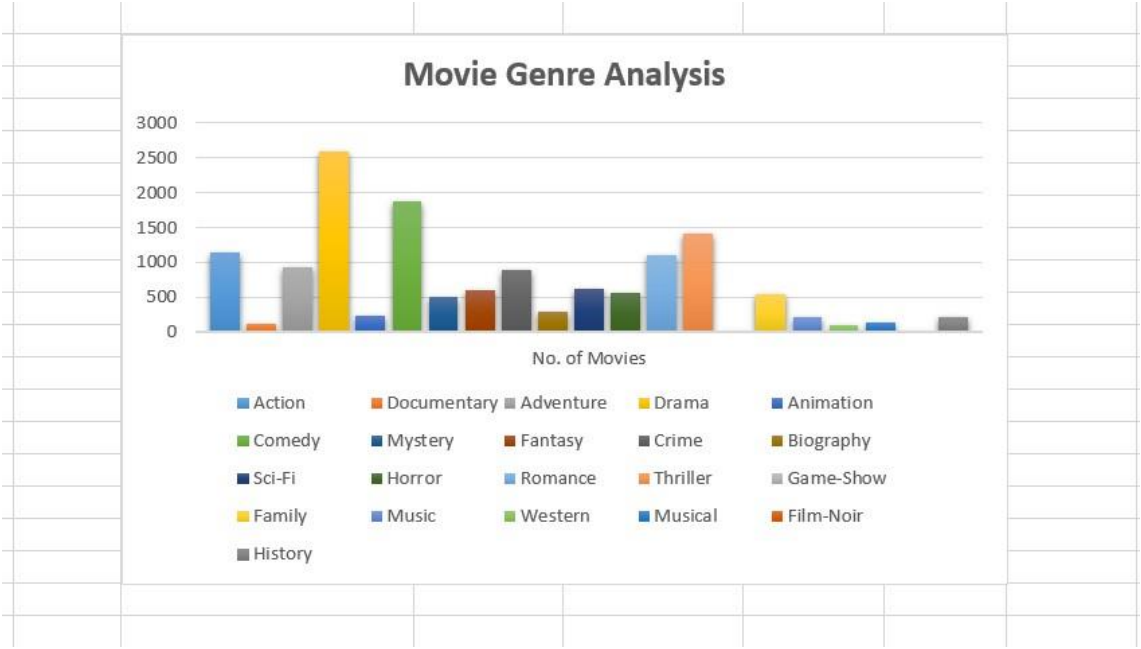
A. Movie Genre Analysis: Analyze the distribution of movie genres and their impact on the IMDB score.

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.

- =COUNTIF(R:X,H3)
- =AVERAGEIF(A2:A5500,H2,C2:C5500)
- =MEDIAN(IF(F:F=H2,C:C))
- =MODE(IF(F:F=H2,C:C))
- =VAR.P(IF(F:F=H2,C:C))
- =MAX(IF(F2:F5044=H2,C2:C5044)) =MIN(IF(F2:F5044=H2,C2:C5044))
- =STDEV.S(IF(F2:F5044=H2,C2:C5044))

genres	No. of Movies	AVERAGE	MEDIAN	MODE	VARIANCE	MAX	MIN	STD DEV
Action	1153	5.972727273	6.3	6.1	1.25070667	9.1	1.6	1.125115866
Documentary	121	6.441848473	7.45	7.5	1.3931335	8.7	1.6	1.187399783
Adventure	923	6.441717913	6.7	7.3	1.24965367	8.6	2.3	1.119115003
Drama	2594	6.441646825	6.9	6.7	0.99304843	9.1	2	0.997031163
Animation	242	6.441238341	6.9	7.1	1.38739049	8.4	3.7	1.187650479
Comedy	1872	6.441107582	6.3	6.4	1.18171231	9.5	1.9	1.08747513
Mystery	500	6.441076037	6.7	6.6	1.24310376	8.5	3.2	1.132232643
Fantasy	610	6.441123908	6.6	7.4	0.79342115	7.9	4.3	0.899265948
Crime	889	6.440854022	7	7.4	0.95112749	9.3	3.1	0.976657883
Biography	293	6.440643623	7.2	7	0.47884527	8.9	4.5	0.693363555
Sci-Fi	616	6.440433141	6.1	6.1	2.04153846	8.2	2.8	1.48716733
Horror	565	6.440341812	5.7	5.9	1.27031139	8.5	2.2	1.129517848
Romance	1106	6.440409461	5.7	5.1	0.63138889	7.1	5.1	0.870440502
Thriller	1408	6.440357853	5.55	5.8	1.61446281	8.1	3.4	1.300516048
Game-Show	1	6.440186916	2.9	0	0	2.9	2.9	0
Family	546	6.44017502	5.7	5.7	3.51900826	8.6	2.8	1.967462602
Music	214	6.440023871	7.2	0	0	7.2	7.2	0
Western	97	6.439992041	7.05	7.1	2.34638889	8.9	3.8	1.5999053
Musical	132	6.439661692	6.7	0	2.375	7.2	3.4	1.779513042
Film-Noir	6	6.442137616	7.6	0	0	7.6	7.6	0
History	207	6.442137616	7.5	0	0	7.5	7.5	0



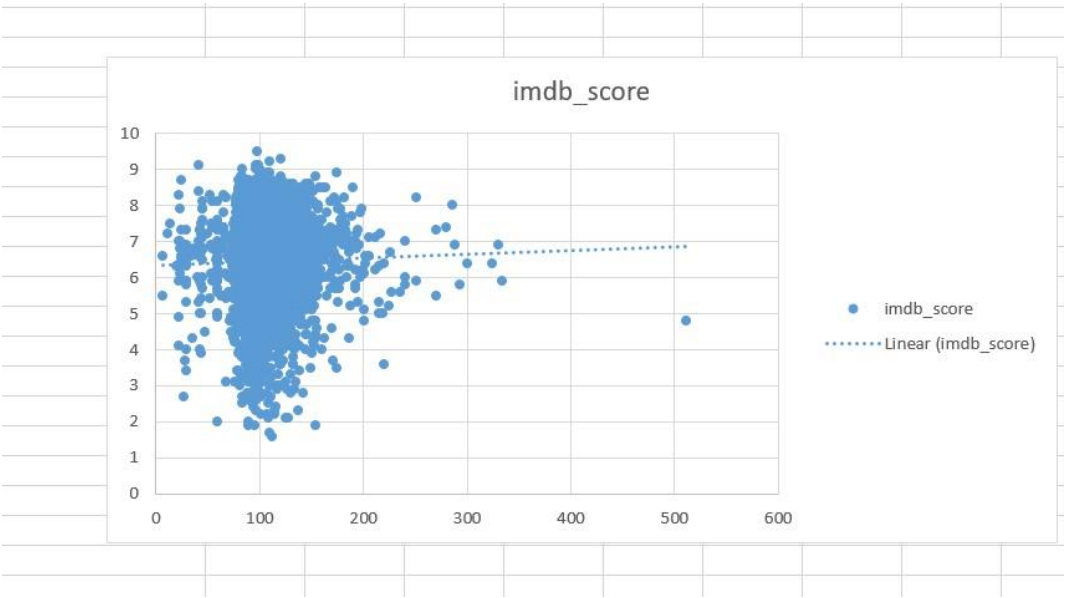
B. Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.

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.

- =AVERAGE(A2:A5500)
- =MEDIAN(A2:A5500)
- =STDEV(A2:A5500)

	E	F
	2. Movie Durations Analysis	
	oprations	**value**
	AVERAGE	107.2011
	MEDIAN	103
	STANDARD DEVIATION	25.19744



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.

=COUNTIF(A2:A5500,E2)

=AVERAGEIF(A2:A5500,E2,C2:C5500)

=MEDIAN(IF(A:A=E2,C:C))

=STDEV.S(IF(A2:A5044=E2,C2:C5044))

D	E	F	G	H	I	J
	language	No. of Movies	Average	MEDIAN	STD DEV	
	English	4704	6.435735544	6.6	1.125716901	
	Japanese	18	6.288888889	6.35	1.465641353	
	French	73	6.539726027	6.7	1.230259653	
	Mandarin	26	6.184615385	6.15	1.179726174	
	Aboriginal	2	7.5	7.5	0.141421356	
	Spanish	40	6.645	6.5	1.085321647	
	Filipino	1	6.5	6.5	0	
	Hindi	28	6.614285714	6.85	1.347346991	
	Russian	11	6.136363636	6.3	1.483423559	
	Maya	1	6.3	6.3	0	
	Kazakh	1	7.3	7.3	0	
	Telugu	1	5.7	5.7	0	
	Cantonese	11	6.4	6.5	0.694262198	
	Icelandic	2	5.65	5.65	1.626345597	
	German	19	6.631578947	6.7	0.797950591	
	Aramaic	1	6.5	6.5	0	
	Italian	11	6.281818182	6.3	1.046726499	
	Dutch	4	6.55	6.45	1.410673598	
	Dari	2	6.2	6.2	0.707106781	
	Hebrew	5	6.46	6.1	0.918150314	
	Chinese	3	7.466666667	7.7	0.776745347	
	Mongolian	1	7.9	7.9	0	
	Swedish	5	6.4	6.4	1.034408043	
	Korean	8	6.6875	6.8	1.004898716	
	Thai	3	6.3	6.4	0.173205081	
	Polish	4	6.9	6.8	0.346410162	
	Bosnian	1	6.5	6.5	0	
	None	2	7.2	7.2	0.565685425	

tor Analysis | E. Budget Analysis

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.

- **Average IMDB score for each director**

  =AVERAGEIF(A:A, A2, B:B)

- **To rank the directors based on their average IMDb scores.**

  =RANK.EQ(C2, $C$2:$C$100, 0)

- **This formula calculates the percentile of each director's IMDb score in comparison to all IMDb scores. Drag this formula down for all directors.**

=PERCENTILE.INC($B$2:$B$100, PERCENTRANK.INC($B$2:$B$100, B2))

	A	B	C	D	E	F	G
1	director_name	imdb_score	AVERAGE	RANK	PERCENTILE	Count Movies	
2	James Cameron	7.9	6.985714	12	7.8	7	
3	Gore Verbinski	7.1	6.885714	18	7.0728	7	
4	Sam Mendes	6.8	6.5375	22	6.7234	8	
5	Christopher Nolan	8.5	7.7625	2	8.4	8	
6	Doug Walker	7.1	7.1	6	7.0728	1	
7	Andrew Stanton	6.6	7.333333	3	6.5706	3	
8	Sam Raimi	6.2	6.492308	23	6.1	13	
9	Nathan Greno	7.8	7.8	1	7.7	1	
10	Joss Whedon	7.5	6.925	15	7.4	4	
11	David Yates	7.5	7.05	7	7.4	4	
12	Zack Snyder	6.9	6.3625	24	6.8746	8	
13	Bryan Singer	6.1	6.925	15	6	8	
14	Marc Forster	6.7	5.8875	29	6.6	8	
15	Gore Verbinski	7.3	6.885714	18	7.2	7	
16	Gore Verbinski	6.5	6.885714	18	6.4152	7	
17	Zack Snyder	7.2	6.3625	24	7.1	8	
18	Andrew Adamson	6.6	6.74	21	6.5706	5	
19	Joss Whedon	8.1	6.925	15	8.09	4	
20	Rob Marshall	6.7	7.28	5	6.6	5	
21	Barry Sonnenfeld	6.8	6.028571	28	6.7234	7	
22	Peter Jackson	7.5	7.05	8	7.4	12	
23	Marc Webb	7	6.966667	14	6.9	3	
24	Ridley Scott	6.7	6.1	27	6.6	17	
25	Peter Jackson	7.9	7.05	8	7.8	12	
26	Chris Weitz	6.1	7	11	6	5	
27	Peter Jackson	7.2	7.05	8	7.1	12	
28	James Cameron	7.7	6.985714	12	7.6	7	
29	Anthony Russo	8.2	7.3	4	8.1	4	

... | IMDB_Movies_Language_Analysis | D. Director Analysis | Director Analysis | E

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.

- ## Correlation

  =CORREL(C2:C5044, B2:B5044)

- **Profit**

  =B2-C2

- ## MAX Function

  = INDEX(E2:E100, MATCH(LARGE(E2:E100, ROW(1:10)), E2:E100, 0))

G	H	I	J
	Top 10 Movies	**Profit**	
	AvatarÂ	523505847	
	King KongÂ	458672302	
	The Chronicles of Narnia: Prince CaspianÂ	403279547	
	X-Men: ApocalypseÂ	358316061	
	InterstellarÂ	265001229	
	Green LanternÂ	214984497	
	Spider-Man 2Â	208992272	
	TangledÂ	208991599	
	Quantum of SolaceÂ	198032628	
	G.I. Joe: The Rise of CobraÂ	181454367	
	SkyfallÂ	173377893	
	How to Train Your DragonÂ	168130696	
	Terminator GenisysÂ	163759914	
	The Sorcerer's ApprenticeÂ	160706665	
	Terminator SalvationÂ	160034110	
	Pacific RimÂ	157358779	
	TitanicÂ	157197282	
	The Jungle BookÂ	142057433	
	Pearl HarborÂ	142000866	

Tech-Stack:

- Microsoft Excel 2022: Utilized for data cleaning, manipulation, and statistical analysis.
- Data Visualization Tools: Incorporated for creating visualizations.

Insights:

1. **Movie Genre Analysis:**
    - Identified the most common genres.
    - Descriptive statistics revealed how different genres impact IMDB scores.
2. **Movie Duration Analysis:**
    - Analyzed the distribution of movie durations.
    - Explored the relationship between movie duration and IMDB scores through descriptive statistics and visualizations.
3. **Language Analysis:**
    - Determined the most common languages in movies.
    - Analyzed the impact of language on IMDB scores.

4. **Director Analysis:**
    - Identified top directors based on average IMDB scores.
    - Analyzed their contribution to movie success using percentile calculations.
5. **Budget Analysis:**
    - Explored the correlation between movie budgets and gross earnings.
    - Identified movies with the highest profit margin.

Results:

1. **Comprehensive Analysis:**
    - Conducted a thorough analysis of various factors influencing movie success on IMDB.
    - Provided detailed insights into the impact of genres, duration, language, directors, and budgets on IMDB scores.
2. **Actionable Insights:**
    - Presented actionable insights for stakeholders in the movie industry.
    - Offered recommendations for optimizing factors leading to higher IMDB ratings.
3. **Improved Decision-Making:**
    - Contributed to the understanding of key success factors in the movie industry.
    - Enabled stakeholders to make informed decisions in future projects, potentially increasing the likelihood of success.

The project, through its detailed analysis and insights, provides a valuable resource for those involved in the movie industry, offering a data-driven approach to decision-making and enhancing the chances of producing successful movies on IMDB.