TRAINITY

IMDB MOVIE ANALYSIS

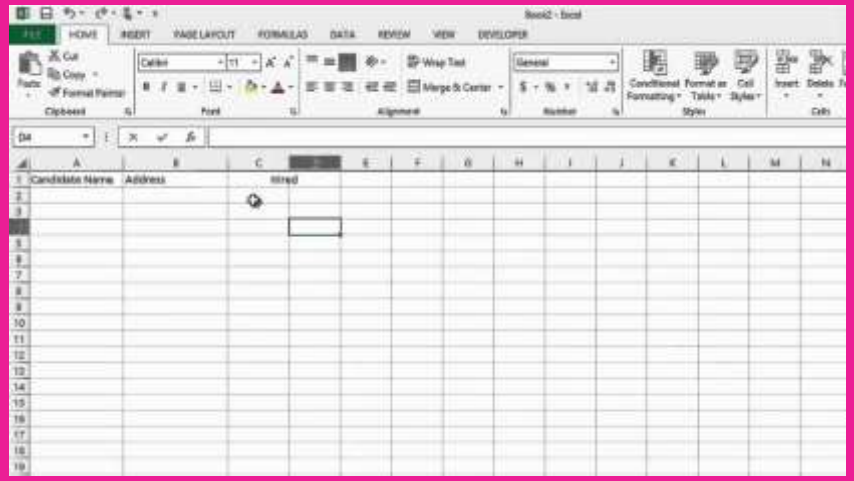
MOGULAGANI PRASHANTH

PROJECT DESCRIPTION

The dataset offered has something to do with IMDB movies. The question of "What factors influence the success of a movie on IMDB?" could be one to look into. High IMDB ratings can be used to measure success in this case. This issue has a big impact on filmmakers, directors, and investors who want to know what makes a movie successful so they can make wise choices for their upcoming projects.

APPROACH

1. *We are using MS excel to solve the problems.*
2. *Microsoft Excel is an application developed by Microsoft for Windows, macOS, Android, and iOS*
3. *We use MS excel formulas to analyze the solutions.*



A.MOVIE GENRE ANALYSIS

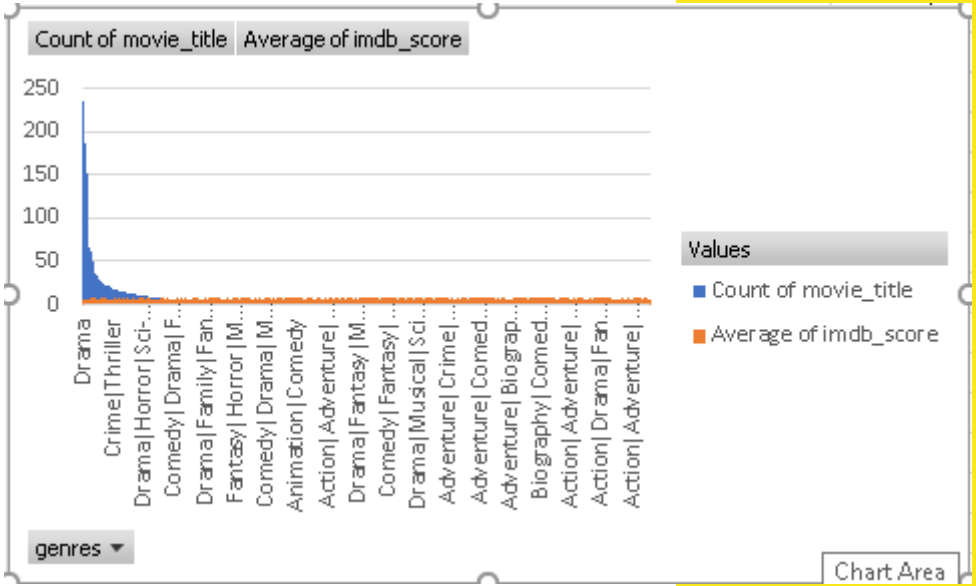
- **Task:** Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

FORMULAS USED **:**

1. To determine the number of genres and the most popular genres, I used a pivot tables.
2. I created a clustered column chart to visually display the data.

IMDB SCORES	Count of movie_title	Average of imdb_score
Drama	235	6.92
Comedy	205	5.83
Comedy|Drama|Romance	187	6.53
Comedy|Drama	187	6.60
Comedy|Romance	156	5.92
Drama|Romance	151	6.94
Crime|Drama|Thriller	99	6.79
Horror	70	5.55
Action|Crime|Drama|Thriller	67	6.45
Drama|Thriller	64	6.42

CLUSTERED COLUMN CHART :



DESCRIPTIVE STATISTICS:

FORMULAS USED:

MEAN:

```
=IF(COUNTIF($A$2:$A$915, A2) = 0, "No Data", AVERAGEIF($A$2:$A$915, A2, $C$2:$C$915))
```

MEDIAN:

```
{=MEDIAN(IF($A$2:$A$915=A2, $B$2:$B$915))}
```

Finding the maximum, minimum, standard deviation, and variance can be done by using pivot tables, adding IMDb scores to the values list, and modifying the value field settings.

Average of imdb_score	MODE	MEAN	MEDIAN	Max of imdb_score	Min of imdb_score	StdDev of imdb_score	Var of imdb_score
6.92	No Mode	6.922553191	235	9.10	3.30	0.93	0.86
5.83	No Mode	5.834146341	205	9.50	1.90	1.29	1.66
6.53	No Mode	6.526737968	187	8.30	4.30	0.80	0.64
6.60	No Mode	6.6	187	8.80	3.30	0.86	0.74
5.92	No Mode	5.917948718	156	8.40	2.70	0.94	0.89
6.94	No Mode	6.935099338	151	8.10	3.50	0.77	0.60
6.79	No Mode	6.787878788	99	9.00	4.80	0.88	0.77
5.55	No Mode	5.551428571	70	8.00	2.20	1.24	1.55
6.45	No Mode	6.447761194	67	9.00	4.60	0.82	0.67
6.42	No Mode	6.415625	64	8.50	3.90	1.02	1.05

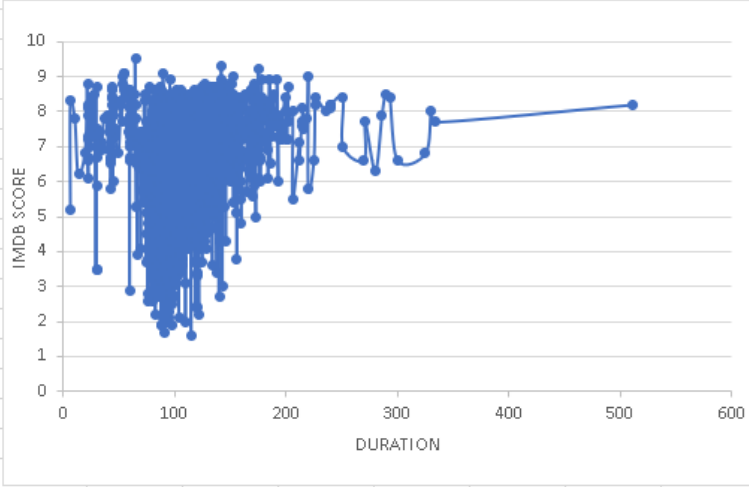
B.MOVIE DURATION ANALYSIS

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

imdb_score	(All)			
DURATION	**FREQUENCY DISTRIBUTION**	**RELATIVE FREQUENCY DISTRIBUTION**	**PERCENT DISTRIBUTION**	**CUMMULATIVE PERCENT**
1-100	2161	0.43	43.38%	43.38%
101-200	2784	0.56	55.89%	99.28%
201-300	32	0.01	0.64%	99.92%
301-400	3	0.00	0.06%	99.98%
501-600	1	0.00	0.02%	100.00%
Grand Total	**4981**	**1.00**	**100.00%**	

I have used PIVOT TABLE to analyze the distribution and I have used filter in pivot table to filter according to imdb score.

duration	imdb_score
511	8.2
334	7.7
330	8
325	6.8
300	6.6
293	8.4
289	8.5
286	7.9
280	6.3
271	7.7
270	6.6
251	7
251	8.4
240	8.2
240	8.1
240	8.2
236	8
227	8.4
226	8.2
225	6.6
220	5.8
220	9



RELATIONSHIP BETWEEN IMDB SCORE AND MOVIE DURATION.

As we can see the relationship between imdb score and duration is the positive relationship.

I have used scatter plot to visualize the relation between movie duration and imdb score.

C.LANGUAGE ANALYSIS

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

Utilizing the count function, the pivot table is used to identify the most prevalent language. The most frequently used langu age in films is English. The IMDB score is entered in the values column and the mean, median, and standard deviation are calculated using the value field settings.

TOP 10

language	Average of imdb_score	StdDevp of imdb_score	MEAN	Count of movie	Count of language	MEDIAN
English	6.398426871	1.121948655	30098.2	4704	4704	6.6
French	7.038356164	0.721989287	513.8	73	73	
Spanish	6.9375	0.844300746	277.5	40	40	
Hindi	6.632142857	1.37374711	185.7	28	28	
Mandarin	6.788461538	1.021810958	176.5	26	26	
German	7.342105263	0.928675225	139.5	19	19	
Japanese	7.394444444	0.962907762	133.1	18	18	
NA	6.85	1.199652728	82.2	12	12	
Russian	6.363636364	1.319278541	70	11	11	
Cantonese	6.954545455	0.671989768	76.5	11	11	

D. DIRECTOR ANALYSIS

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

THESE ARE THE TOP 10 DIRECTORS BASED ON THE IMDB SCORE.

director_name	imdb_score	average imdb scores	pecentrank
John Blanchard	9.5	9.5	100.0%
Mitchell Altieri	8.7	8.7	99.9%
Sadyk Sher-Niyaz	8.7	8.7	100.0%
Cary Bell	8.7	8.7	100.0%
Mike Mayhall	8.6	8.6	100.0%
Charles Chaplin	8.6	8.6	100.0%
Raja Menon	8.5	8.5	99.9%
Ron Fricke	8.5	8.5	99.9%
Damien Chazelle	8.5	8.5	100.0%
Majid Majidi	8.5	8.5	100.0%

Using a pivot table to find the top directors based on average IMDB scores.

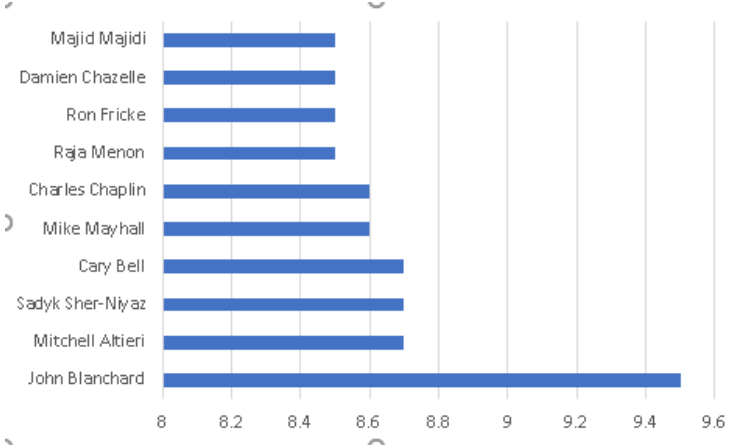
90TH % PERCENTILE	7.55

FORMULAS USED:

PERCENTILE: `=PERCENTILE($C$2:$C$5044, 0.9)`

PERCENTRANK: `=PERCENTRANK.EXC($C2:$C5044,C2,4)`

CLUSTERED BAR:



E.BUDGET ANALYSIS

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

budget	gross	CORRELATION
237000000	760505847	0.102179454
300000000	309404152	
245000000	200074175	
250000000	448130642	
NA	NA	
263700000	73058679	
258000000	336530303	
260000000	200807262	
250000000	458991599	
250000000	301956980	
250000000	330249062	
209000000	200069408	
200000000	168368427	
225000000	423032628	
215000000	89289910	

CORRELATION BETWEEN MOVIE BUDGETS AND GROSS EARNINGS: **0.102179454**

FORMULA USED :

=CORREL(I:I,H:H)

PROFIT	movie_title
523505847	Avatar
9404152	Pirates of the Caribbean: At World's End
198130642	The Dark Knight Rises
78530303	Spider-Man 3
208991599	Avengers: Age of Ultron
51956980	Harry Potter and the Half-Blood Prince
80249062	Batman v Superman: Dawn of Justice
198032628	Pirates of the Caribbean: Dead Man's Chest
66021565	Man of Steel
403279547	The Avengers

FORMULA USED :

GROSS-BUDGET

THESE ARE THE TOP 10 MOVIES HAVING HIGHEST PROFIT MARGIN FROM THE GIVEN DATASET

RESULT

I became acquainted with new EXCEL features, lingo, and methods.
By obtaining the appropriate insights from the problem description, practical problems can be resolved. Thanks to the concepts, I was able to comprehend the description of the problem. This project has improved my problem-solving skills and taught me how to apply the theoretical concepts I learned in training to actual-world circumstances.

LINK OF EXCEL FILE :

https://docs.google.com/spreadsheets/d/1eCHjsGwN7sG6QRUPUsLfJGj6MChYn0G_/edit?usp=drive_link&ouid=108547673521600619650&rtpof=true&sd=true