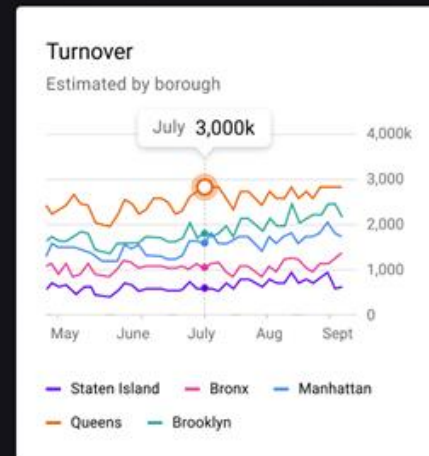


IMDB MOVIE ANALYSIS

FINAL PROJECT-1



By - Anindya Das

DESCRIPTION

Problem Statement: The provided dataset relates to IMDB movies, and our goal is to explore the question: "What factors influence a movie's success on IMDB?" In this context, we define success as achieving high IMDB ratings. Analyzing this issue is vital for movie producers, directors, and investors, as it will help them identify the elements that contribute to a movie's success, enabling them to make well-informed decisions for their future projects.

APPROACH

Data cleaning: this step involves preparing the data for analysis by addressing missing values, eliminating duplicates, converting data types if needed, and potentially performing feature engineering.

Data Analysis: In this section, analyze the data to understand the relationships between variables like genre, director, budget, year of release, and actors in relation to movie ratings.

TECH-STACK USED

The primary tech stack for this project will be excel, which provides a comprehensive range of functions for data analysis and manipulation. It is ideal for exploratory data analysis and insights generation. Key features like formulas, pivot tables, and charts will be used for data cleaning, calculating statistics, identifying trends, and visualizing hiring data, all through a user-friendly interface.



Microsoft Excel

DATA ANALYTICS TASKS:

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.

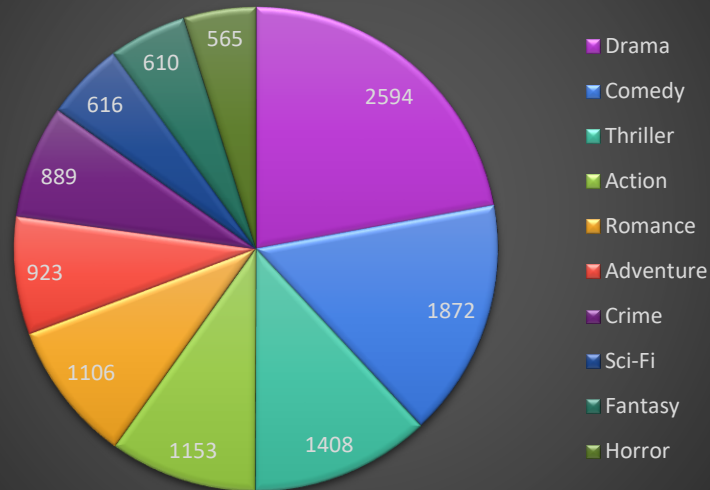
LIST_OF_GENRES	COUNT_OF_GENRES
Adventure	923
Thriller	1408
Animation	242
Family	546
Fantasy	610
Romance	1106
Crime	889
Comedy	1872
Drama	2594
Sci-Fi	616
Horror	565
Mystery	500
Western	97
History	207
Musical	132
Music	214
War	213
Biography	293
Sport	182
Reality-TV	2
Documentary	121
Short	5
Film-Noir	6
News	3
Action	1153
Total	14499

DESCRIPTIVE STATISTICS OF IMDB SCORE

Descriptive statistics of IMDB Score

Mean	6.442137616
Median	6.6
Mode	6.7
Max	9.5
Min	1.6
Range	7.9
Variance	1.265885711
Standard Deviation	1.125115866

Top 10 Genres



Top 10 Genres

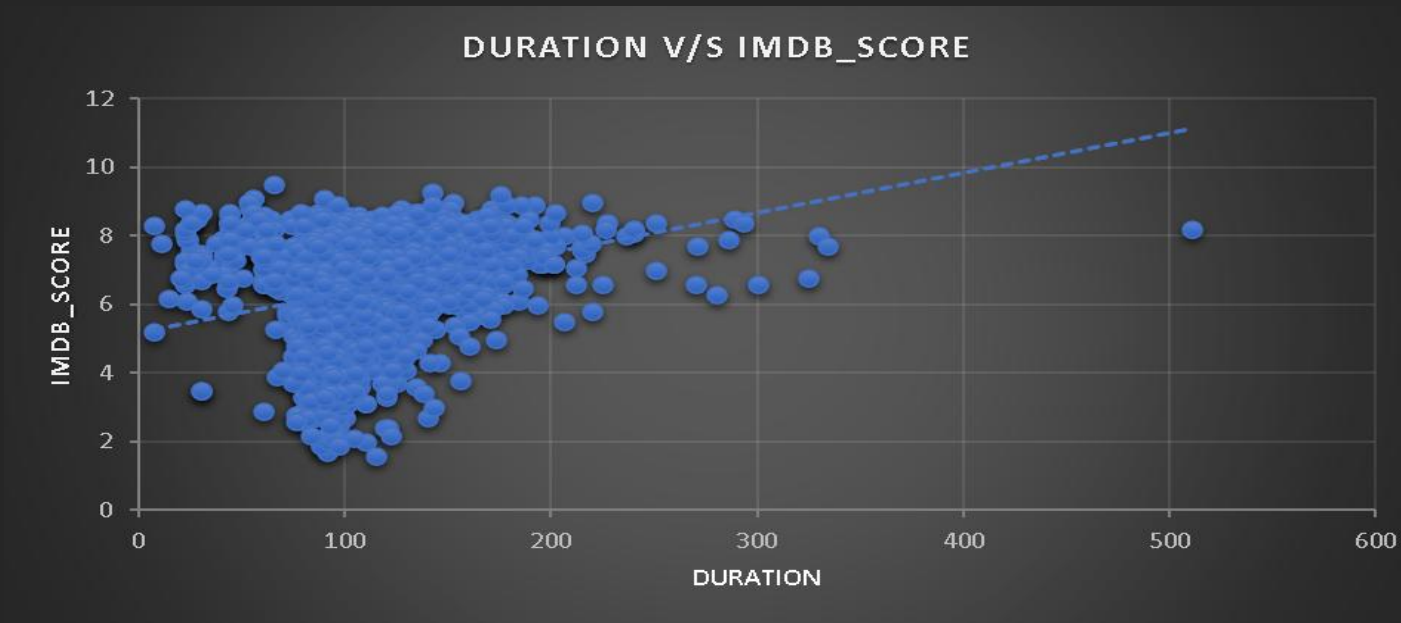
Row Labels	Sum of COUNT_OF_GENRES
Drama	2594
Comedy	1872
Thriller	1408
Action	1153
Romance	1106
Adventure	923
Crime	889
Sci-Fi	616
Fantasy	610
Horror	565
Grand Total	11736

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.

Descriptive Statistics of Movie Duration	
Mean	107.201074
Median	103
Standard Deviation	25.19744081

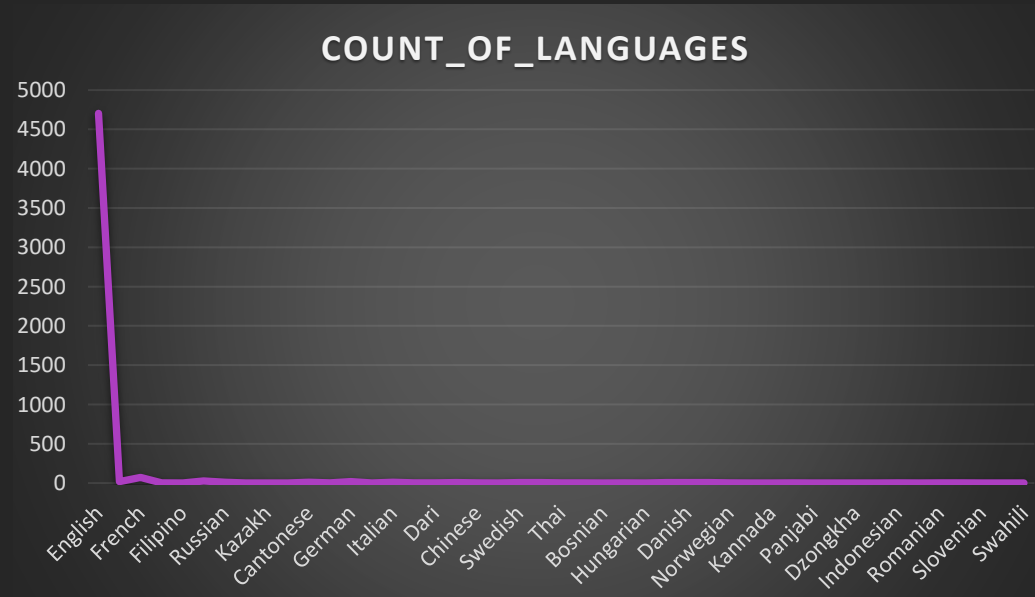


LANGUAGE ANALYSIS:

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.

Descriptive statistics of IMDB Score	
MEAN	6.441164778
MEDIAN	6.6
STANDARD DEVIA	1.12475531



language	COUNT_OF_LANGUAGES
English	4704
Japanese	18
French	73
Aborigin	2
Filipino	1
Hindi	28
Russian	11
Maya	1
Kazakh	1
Telugu	1
Cantone	11
Icelandic	2
German	19
Aramaic	1
Italian	11
Dutch	4
Dari	2
Hebrew	5
Chinese	3
Mongolia	1
Swedish	5
Korean	8
Thai	3
Polish	4
Bosnian	1
None	2
Hungaria	1
Portugue	8
Danish	5
Arabic	5
Norwegia	4
Czech	1
Kannada	1
Zulu	2
Panjabi	1
Tamil	1
Dzongkh	1
Vietnam	1
Indonesi	2
Urdu	1
Romania	2
Persian	4
Slovenia	1
Greek	1
Swahili	1
Total	4965

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.

90% MOVIES SCORE	NO OF MOVIES>7.8
7.8	403

Directors	Average of imdb_score
John Blanchard	9.5
Sadyk Sher-Niyaz	8.7
Mitchell Altieri	8.7
Cary Bell	8.7
Mike Mayhall	8.6
Charles Chaplin	8.6
Damien Chazelle	8.5
Ron Fricke	8.5
Raja Menon	8.5
Majid Majidi	8.5
Grand Total	8.68

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.

HIGHEST PROFIT MARGIN	CORRELATION
523505847	0.099347223

D2				
		X	✓	fx
				=(B:B-C:C)
	A	B	C	D
1	movie_title	gross	budget	PROFIT MARGIN
2	Avatar	760505847	237000000	523505847
3	Pirates of the Caribbean: At World's End	309404152	300000000	9404152
4	Spectre	200074175	245000000	-44925825
5	The Dark Knight Rises	448130642	250000000	198130642
6	John Carter	73058679	263700000	-190641321
7	Spider-Man 3	336530303	258000000	78530303
8	Tangled	200807262	260000000	-59192738
9	Avengers: Age of Ultron	458991599	250000000	208991599
10	Harry Potter and the Half-Blood Prince	301956980	250000000	51956980
11	Batman v Superman: Dawn of Justice	330249062	250000000	80249062
12	Superman Returns	200069408	209000000	-8930592
13	Quantum of Solace	168368427	200000000	-31631573
14	Pirates of the Caribbean: Dead Man's Chest	423032628	225000000	198032628
15	The Lone Ranger	89289910	215000000	-125710090
16	Man of Steel	291021565	225000000	66021565
17	The Chronicles of Narnia: Prince Caspian	141614023	225000000	-83385977
18	The Avengers	623279547	220000000	403279547
19	Pirates of the Caribbean: On Stranger Tides	241063875	250000000	-8936125
20	Men in Black 3	179020854	225000000	-45979146
21	The Hobbit: The Battle of the Five Armies	255108370	250000000	5108370
22	The Amazing Spider-Man	262030663	230000000	32030663
23	Robin Hood	105219735	200000000	-94780265
24	The Hobbit: The Desolation of Smaug	258355354	225000000	33355354
25	The Golden Compass	70083519	180000000	-109916481
26	King Kong	218051260	207000000	11051260
27	Titanic	658672302	200000000	458672302
28	Captain America: Civil War	407197282	250000000	157197282

INSIGHTS

1. The Most common movie genres from the dataset are Drama, Comedy, Thriller, and Action.
2. The average duration of a movie is 107 minutes.
3. The Most common languages used in the movies are English, French, Spanish, Mandarin, and German.
4. I have identified that John Blanchard, Sadyk Sher-Niyaz, Mitchell Altieri, Cary Bell, Mike Mayhall, Charles Chaplin, Damien Chazelle, Ron Fricke, Raja Menon, and Majid Majidi are the top 10 directors with average IMDb score ≥ 8.5 .
5. The Top with the highest profits are Avatar, Jurassic World, Titanic, Star Wars: Episode IV – A New Hope, and E.T. The Extra-Terrestrial. The Correlation between budget and gross is positive.

CONCLUSION

Dataset with Analysis:

https://docs.google.com/spreadsheets/d/1ORPF144yoAaxHSUefqpj5ZPB0h-lit2F/edit?usp=drive_link&oid=100865564169059724511&rtpof=true&sd=true

The background is a dark gray color. It features several faint, light gray geometric patterns. In the top right corner, there is a large circular motif with concentric circles and radial lines, resembling a compass or a stylized sun. In the bottom right corner, there is a smaller circular motif with concentric circles and a dashed line. In the bottom left corner, there is another circular motif with concentric circles and a dashed line. The text "Thank you!" is centered in the middle of the image in a bold, yellow, sans-serif font.

Thank you!