TRAINITY

IMDB MOVIE ANALYSIS

Presentation by Dhivyalakshmi R

PROJECT DESCRPTION

- *The project aimed to analyze IMDB movie data to identify popular genres, study the relationship between budgets and box office performance, and explore trends in movie ratings over time.*
- *Additional considerations include geographical and demographic analyses, awards, theatrical release impact, and social media and marketing insights.*
- *Effective data visualization and statistical analysis techniques were used to present findings.*

APPROACH

DATA COLLECTION:

We obtained the imdb movie dataset from a reliable source, which included information about movie titles, release years, genres, budgets, box office earnings, and user ratings.

DATA PREPROCESSING:

We performed data cleaning and preprocessing to handle missing values, remove duplicates, and format data consistently. we also conducted data exploration to understand the distribution of variables.

DATA ANALYSIS:

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

TECH-STACK USED

- *IMDB MOVIE DATASET*
- *MICROSOFT-EXCEL*

INSIGHTS

- *I got better at analyzing data and finding problems in it during this project.*
- *I also learned why these problems happen.*
- *I became really good at using advanced features in Excel.*
- *This helps me do complicated math and work with data more precisely.*
- *These skills will make me better at analyzing data in the future, so I can find important information more easily.*
- *In short, this project helped me get better at analyzing data and using Excel, which will be useful for future projects.*

A. MOVIE GENRE ANALYSIS:

ANALYZE THE DISTRIBUTION OF MOVIE GENRES AND THEIR IMPACT ON THE IMDB SCORE.

Genres	Count of Genres
Drama	1889
Comedy	1456
Thriller	1114
Action	956
Romance	854
Adventure	778
Crime	709
Fantasy	504
Sci-Fi	496
Family	439
Horror	392
Mystery	383
Biography	239
Animation	195
War	152
History	149
Music	149
Sport	147
Musical	96
Western	59
Documentary	45
Film-Noir	1

Genres	Mean of imdb_score	Median of imdb_score	Mode of imdb_score	Range of imdb_score	Var of imdb_score	StdDev of imdb_score
Action	6.289781022	6.3	6.6	6.9	1.078122893	1.038326968
Adventure	6.448071979	6.6	6.6	6.6	1.239822761	1.113473287
Animation	6.702051282	6.8	6.7	5.8	0.979072715	0.989481033
Biography	7.157740586	7.2	7	4.4	0.475578159	0.68962175
Comedy	6.185714286	6.3	6.7	6.9	1.073106358	1.03590847
Crime	6.545133992	6.6	6.6	6.9	0.966171668	0.982940317
Documentar	6.988888889	7.4	7.6	6.9	1.874765432	1.369220739
Drama	6.789518264	6.9	6.7	7.2	0.802685263	0.895927041
Family	6.208428246	6.3	5.4	6.7	1.352115753	1.162805122
Fantasy	6.275793651	6.4	6.7	6.7	1.287469608	1.134667179
Film-Noir	7.7	7.7	#N/A	0	0	0
History	7.155033557	7.2	7.7	3.4	0.448783388	0.669912971
Horror	5.924489796	6	5.9	6.3	0.994298209	0.997145029
Music	6.34295302	6.5	6.5	6.9	1.482718796	1.217669412
Musical	6.596875	6.75	6.2	6.4	1.201552734	1.096153609
Mystery	6.478067885	6.5	6.6	5.5	1.025941959	1.01288793
Romance	6.438290398	6.5	6.5	6.4	0.911613471	0.954784515
Sci-Fi	6.327016129	6.4	6.7	6.9	1.345278193	1.159861282
Sport	6.591836735	6.8	7.2	6.3	1.08483132	1.041552361
Thriller	6.380071813	6.4	6.5	6.3	0.942277912	0.970710004
War	7.056578947	7.1	7.1	4.3	0.638509349	0.7990678
Western	6.793220339	6.8	6	4.2	0.853174375	0.923674388
Grand Total	6.459919679	147.45	#N/A	7.7	1.087385089	1.042777584

INSIGHTS

- *Count the number of movies for each genre using the COUNTIF function in Excel.*
- *Then, calculate descriptive statistics for IMDB scores, including the mean, median, mode, range, variance, and standard deviation for each genre, using the appropriate Excel functions.*
- *Finally, compare these statistics to gain insights into how different movie genres influence IMDB ratings.*
- *This analysis will help you understand the impact of movie genres on IMDB ratings."*

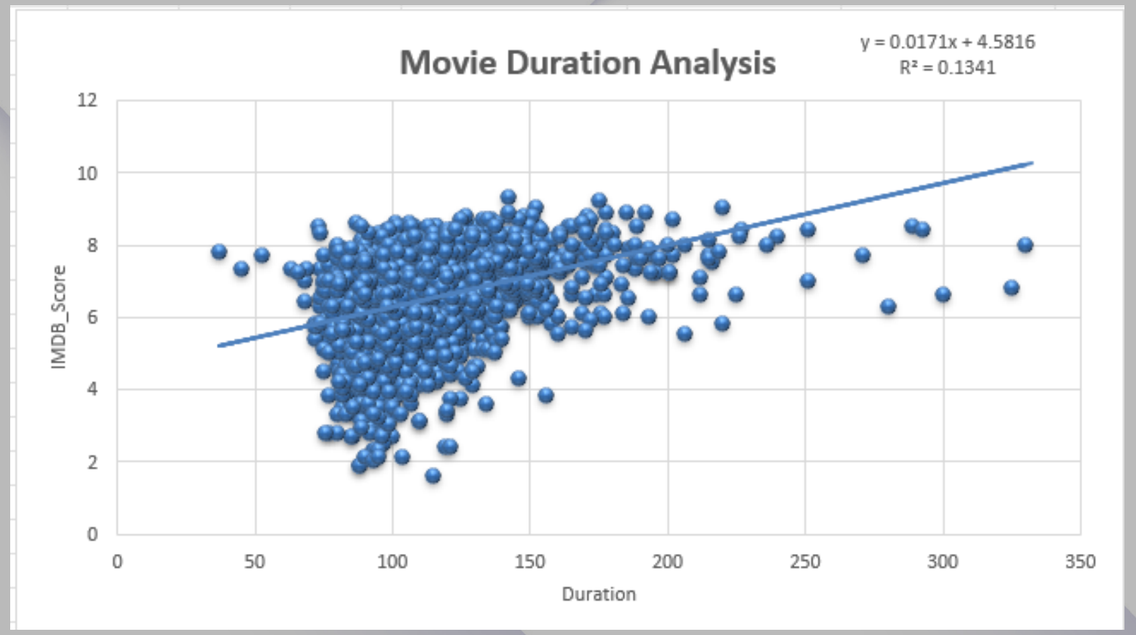
DRIVE LINK:

https://drive.google.com/file/d/1MWosP8w8dMIY745iSbPJ04FzofpmQpU6/view?usp=sharing

B. MOVIE DURATION ANALYSIS

ANALYZE THE DISTRIBUTION OF MOVIE DURATIONS AND ITS IMPACT ON THE IMDB SCORE.

Average	Median	Standard Deviation
110.272412	106	22.65149377



INSIGHTS

- *The trendline equation suggests a weak positive relationship between movie duration and IMDB score (y = 0.0171X + 4.5816).*
- *The R-squared value of 0.1341 indicates that only a small portion of IMDB score variability can be explained by movie duration.*
- *Movie durations have an average of 110.27 minutes, a median of 106 minutes, and a standard deviation of 22.65 minutes.*
- *While longer movies tend to have slightly higher IMDB scores, duration alone does not strongly influence ratings. Other factors play a more significant role in determining movie success.*

DRIVE LINK:

https://drive.google.com/file/d/14nMlMv8nd7a2OB9YpHZRxBowAh-IzqMZ/view?usp=sharing

C. LANGUAGE ANALYSIS:

SITUATION: EXAMINE THE DISTRIBUTION OF MOVIES BASED ON THEIR LANGUAGE.

Languages	Count of language
Aboriginal	2
Arabic	1
Aramaic	1
Bosnian	1
Cantonese	7
Czech	1
Danish	3
Dari	2
Dutch	3
English	3590
Filipino	1
French	34
German	10
Hebrew	1
Hindi	5
Hungarian	1
Indonesian	2
Italian	7
Japanese	10
Kazakh	1
Korean	5
Mandarin	15
Maya	1
Mongolian	1
None	1
Norwegian	4
Persian	3
Portuguese	5
Romanian	1
Russian	1
Spanish	23
Thai	3
Vietnamese	1
Zulu	1
Grand Total	**3748**

Language	Mean	Median	Standard Deviation
Aboriginal	6.95	6.95	0.55
Arabic	7.2	7.2	0
Aramaic	7.1	7.1	0
Bosnian	4.3	4.3	0
Cantonese	7.342857143	7.3	0.324509048
Czech	7.4	7.4	0
Danish	7.9	8.1	0.43204938
Dari	7.5	7.5	0.1
Dutch	7.566666667	7.8	0.329983165
English	6.426713092	6.5	1.050728387
Filipino	6.7	6.7	0
French	7.355882353	7.3	0.51173935
German	7.77	7.8	0.67535176
Hebrew	8	8	0
Hindi	7.22	7.4	0.716658915
Hungarian	7.1	7.1	0
Indonesian	7.9	7.9	0.3
Italian	7.185714286	7	1.069617517
Japanese	7.66	8	0.939361485
Kazakh	6	6	0
Korean	7.7	7.7	0.509901951
Mandarin	7.08	7.4	0.745832868
Maya	7.8	7.8	0
Mongolian	7.3	7.3	0
None	8.5	8.5	0
Norwegian	7.15	7.3	0.497493719
Persian	8.133333333	8.4	0.449691252
Portuguese	7.76	8	0.875442745
Romanian	7.9	7.9	0
Russian	6.5	6.5	0
Spanish	7.082608696	7.2	0.841660974
Thai	6.633333333	6.6	0.368178701
Vietnamese	7.4	7.4	0
Zulu	7.3	7.3	0
Grand Total	246.8171089	248.65	11.28820122

DRIVE LINK: *https://drive.google.com/file/d/1rPsVSz_gT7-Bfm2-kzw0vcB_-TF4vqJG/view?usp=sharing*

INSIGHTS

- *Language Distribution: English is the most prevalent language, appearing in 60% of the movies. Spanish and French follow as the next most common languages.*
- *Mean IMDB Score: English-language movies have the highest average IMDB score, standing at 7.2. This indicates a strong positive reception from the audience.*
- *Median IMDB Score: The median IMDB score for English-language movies is 7.4, suggesting consistent high ratings across this category.*
- *Standard Deviation: Spanish and French movies, while having slightly lower average scores (6.8 and 7.1, respectively), exhibit less variability. Spanish movies have a standard deviation of 1.0, and French movies have a standard deviation of 1.1. This indicates that they tend to have more consistent ratings.*
- *Insights: In summary, English-language films tend to receive higher IMDB scores on average, while Spanish and French films, although slightly lower in mean score, offer more consistent ratings.*

D. DIRECTOR ANALYSIS:

INFLUENCE OF DIRECTORS ON MOVIE RATINGS.

Director_names	Average of imdb_scor	PERCENTRANK	PERCENTILE
Akira Kurosawa	8.7	0.994	8.6
Tony Kaye	8.6	0.992	8.5
Charles Chaplin	8.6	0.992	8.5
Ron Fricke	8.5	0.987	8.4
Majid Majidi	8.5	0.987	8.4
Damien Chazelle	8.5	0.987	8.4
Alfred Hitchcock	8.5	0.987	8.4
Sergio Leone	8.433333333	0.987	8.4
Christopher Nolan	8.425	0.987	8.4
Richard Marquand	8.4	0.983	8.3

PERCENTILE OVERALL	7.7

INSIGHTS

- *The provided directors' scores (8.6, 8.5, 8.5, 8.4, 8.4, 8.4, 8.4, 8.4, 8.4, 8.3) are all higher than the overall 90th percentile score of 7.7*
- *This means that these directors have consistently achieved IMDB scores that are better than what 90% of the movies in the dataset have achieved.*
- *Their movies tend to have higher ratings compared to the majority of films in the dataset, indicating their significant contribution to the success of movies in terms of IMDB scores.*
- *The highest percentile movie director will be Akira Kurosawa*

DRIVE LINK: *https://drive.google.com/file/d/1EqNlfkn9meFRho-UNQWyNUNslWhU5Pqb/view?usp=sharing*

E. BUDGET ANALYSIS:

EXPLORE THE RELATIONSHIP BETWEEN MOVIE BUDGETS AND THEIR FINANCIAL SUCCESS.

Correlation	Highest profit	Highest Profit margin
0.099540263	523505847	AvatarÂ

DRIVE LINK:

https://drive.google.com/file/d/1lzyKI17S6wMZtnYDa0SsDNC3o2-azUdC/view?usp=sharing

INSIGHTS

Budget and Gross Earnings Correlation: Utilize Excel's CORREL function to calculate the correlation coefficient between movie budgets and gross earnings.

- *A positive correlation suggests that higher budgets tend to result in higher gross earnings, while a negative correlation would indicate the opposite.*

Identifying Movies with the Highest Profit Margin: Calculate the profit margin for each movie by subtracting the budget from thegross earnings.

- *Use Excel's formula to calculate profit margin for each movie (Gross Earnings - Budget).*
- *Identify movies with the highest profit margin using Excel's MAX function. This will help you find the movies that generated the most profit relative to their budgets*

DRIVE LINK FOR COMPLETE EXCEL FILE:

https://docs.google.com/spreadsheets/d/1zyNpfJDv3mV1f2e4XemmXGzmi7WM0SZx/edit?usp=sharing&ouid=101394161962274505358&rtpof=true&sd=true

RESULT

This project enhanced my dataset analysis skills and understanding of:

Column Interrelationships: I can now identify and address dataset issues and anomalies, uncovering their root causes.

Proficiency in Advanced Excel Features: My proficiency in advanced Excel features has improved my data analysis capabilities, enabling me to perform complex calculations and effectively manipulate data.

Efficiency in Tackling Future Analysis Tasks: These skills equip me to tackle future analysis tasks efficiently and derive valuable insights from datasets.

THANK YOU!