

IMDB MOVIE ANALYSIS



Submitted By : Shubham Saxena

Project Description

IMDb (an acronym for Internet Movie Database) is an online database of information related to films, television series, podcasts, home videos, video games, and streaming content online – including cast, production crew and personal biographies, plot summaries, trivia, ratings, and fan and critical reviews. Here we're provided with IMDb's dataset for movies from 1920-2010, which contains information about movies, its actors, directors, budget, collection, rating etc. We'll clean the dataset and get answer for asked questions by using Five why methods for analytics using office 365 excel.

Approach

For this project, first we'll get understanding of the given data. Then we'll clean the data as per our requirement by removing null values, delete unnecessary columns, etc. After the cleaning we'll using pivot table, various functions and charts for desired answers for the questions. We'll continues ask why to data to get in depth of the root of the problem. At the end we'll present our answers with proper formatting in table and graphs.

Five “WHY” Approach

WHY 1

"Why do movies with higher budgets tend to have higher ratings?"

WHY 2

"Why does better production quality lead to higher ratings?"

WHY 3

"Why does an enhanced viewer experience lead to higher ratings?"

Five “WHY” Approach

WHY 4

"Why are viewers more likely to rate a movie highly if they enjoyed watching it?"

WHY 5

"Why do positive reviews matter?"

Tech-Stack Used:

- 1. MS Office Word**
- 2. MS Office Excel**
- 3. Google Sheet**

INSIGHTS

A. Cleaning The Data

This is one of the most important steps to perform before moving forward with the analysis.

Cleaning can be done by:

- Dropping unnecessary 14 columns like Color, Director_facebook_likes, actor_3_facebook_likes, actor_2_name, actor_1_facebook_likes, casr_total_facebook_likes, actor_3_name, facenumber_in_posts, plot_keywords, movie_imdb_link, Content_rating, actor_2_facebook_likes, aspect_ratio, and Movie_facebook_likes. These columns should be removed as these don't have any important role with the tasks which we need to perform.
- Find Blank Cell/Null Value by Go To special and delete them by rows.
- 135 Duplicates found and deleted by using remove duplicates option from data tab

Task 1

Movie Genre Analysis

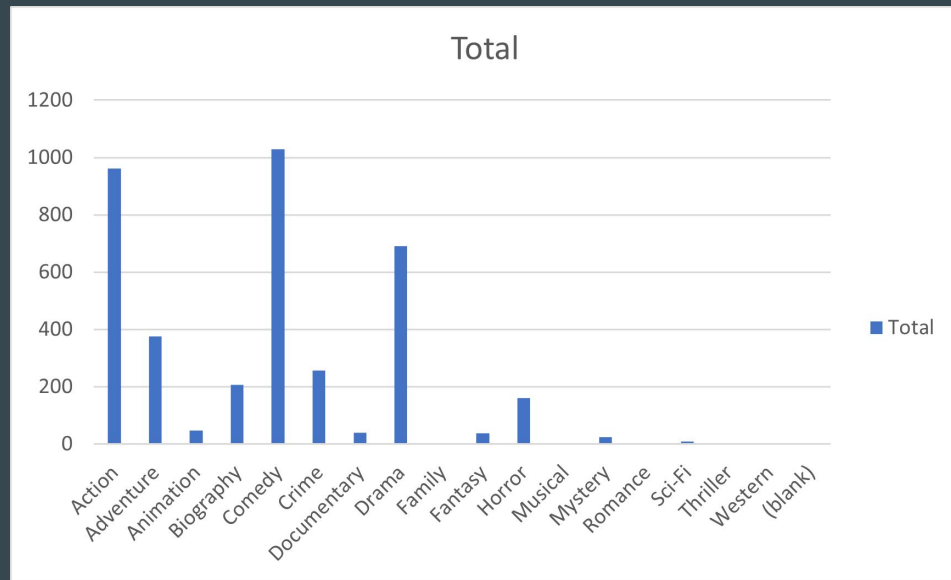
Analyse the distribution of movie genres and their impact on the IMDB score.

- 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.

Task 1 Analysis

Findings

Average	226.3529412
Median	40
Mode	3
Max	1029
Min	2
Range (Max-Min)	1027
Variance	116974.7426
Standard Deviation	342.0157053



[VIEW SOURCE FILE](#)

Task 2

Movie Duration Analysis

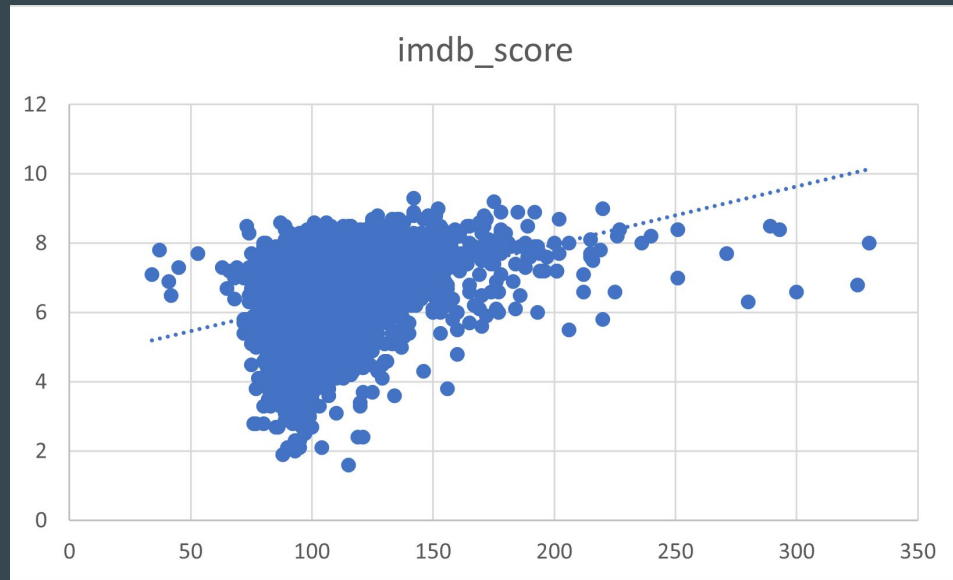
Analyze the distribution of movie durations and its impact on the IMDB score.

- 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.

Task 2 Analysis

Findings

Average Duration	109.9241
Median	106
Standard Deviation	22.75365



[VIEW SOURCE FILE](#)

Task 3

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.

Task 3 Analysis

Findings

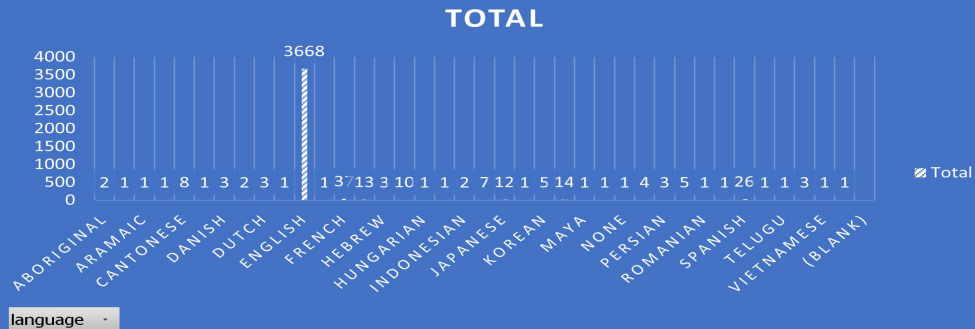
[VIEW SOURCE FILE](#)

1. Count of Genres using countif function

Language	Count Of Genres	Mean	Median	Standard Deviaton
Aboriginal	2	6.95	6.5	1.047657909
Arabic	1	7.2	7.2	NA
Aramaic	1	7.1	7.1	NA
Bosnian	1	4.3	4.3	NA
Cantonese	8	7.2375	6.6	1.10513431
Czech	1	7.4	7.4	NA
Danish	3	7.9	6.8	1.113778941
Dari	2	7.5	6.7	1.114677441
Dutch	3	7.566667	6.6	1.13585388
Dzongkha	1	7.5	7.5	NA
English	3668	6.423909	6.6	1.053794573
Filipino	1	6.7	6.7	NA
French	37	7.286486	6.5	1.061380251
German	13	7.692308	6.7	1.1127341
Hebrew	3	7.5	6.8	0.983011796
Hindi	10	6.76	6.7	1.092956828
Hungarian	1	7.1	7.1	NA
Icelandic	1	6.9	6.9	NA
Indonesian	2	7.9	6.8	1.10691927
Italian	7	7.185714	6.6	1.113545399
Japanese	12	7.625	6.6	1.096013453
Kazakh	1	6	6	NA
Korean	5	7.7	6.7	1.097205546
Mandarin	14	7.021429	6.5	1.049310634

Maya	1	7.8	7.8	NA
Mongolian	1	7.3	7.3	NA
None	1	8.5	8.5	NA
Norwegian	4	7.15	6.8	1.091486743
Persian	3	8.133333	6.8	0.894089557
Portuguese	5	7.76	6.8	1.116081142
Romanian	1	7.9	7.9	NA
Russian	1	6.5	6.5	NA
Spanish	26	7.05	6.5	1.061539123
Swedish	1	7.6	7.6	NA
Telugu	1	8.4	8.4	NA
Thai	3	6.633333	6.7	1.08457693
Vietnamese	1	7.4	7.4	NA
Zulu	1	7.3	7.3	NA
Grand Total	3848			

Count of Movie Name



Task 4

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.

Task 4 Analysis

Findings

Top 10 Directors	Average of IMDB Score
Charles Chaplin	8.6
Tony Kaye	8.6
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Ron Fricke	8.5

Sergio Leone	8.433333333
Christopher Nolan	8.425
Asghar Farhadi	8.4
Marius A. Markevicius	8.4

Large IMDB Score	8.6
Percentage	0.888
Percentile	8.5992

[VIEW SOURCE FILE](#)

Task 5

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 MAX function.

Task 5 Analysis

Findings

Correlation	0.100850218
Highest Profit Margin	523505847

[VIEW SOURCE FILE](#)

LINKS

MS OFFICE WORD FILE

[Click Here](#)

MS OFFICE EXCEL

[Click Here](#)

THANK YOU
