IMDB Movie Analysis

Here’s where presentation begins

By Anindya Das

Table of contents

01

Introduction

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

02

Presentation & Analysis

Summarize the insights and knowledge gained during the project. Discuss key findings and any meaningful trends or patterns discovered.

03

Conclusion

Describe what you achieved through the project and how it has contributed to your understanding of the IMDB Movie Analysis

01
Introduction

"What factors influence the success of a movie on IMDB?"

Detail Analysis

Here's a brief description of each column:
color: This column might indicate whether the movie is in color or black and white.
director_name: The name of the movie's director.
num_critic_for_reviews: The number of critic reviews for the movie.
duration: The duration (in minutes) of the movie.
director_facebook_likes: The number of Facebook likes the director has.
actor_3_facebook_likes: The number of Facebook likes the third actor (possibly a supporting role) has.
actor_2_name: The name of the second actor (possibly a supporting role).
actor_1_facebook_likes: The number of Facebook likes the first actor (possibly the lead actor) has.
gross: The gross earnings (box office revenue) of the movie.
genres: The genre(s) of the movie.
actor_1_name: The name of the first actor (possibly the lead actor).
movie_title: The title of the movie.
num_voted_users: The number of users who voted for the movie.
cast_total_facebook_likes: The total number of Facebook likes for the movie's cast.
actor_3_name: The name of the third actor (possibly a supporting role).
facenumber_in_poster: The number of faces on the movie poster.
plot_keywords: Keywords or phrases related to the movie's plot.
movie_imdb_link: The IMDb link to the movie.
num_user_for_reviews: The number of user reviews for the movie.
language: The language of the movie.
country: The country where the movie was produced.
content_rating: The content rating of the movie (e.g., PG-13, R, etc.).
budget: The budget of the movie.
title_year: The year the movie was released.
actor_2_facebook_likes: The number of Facebook likes the second actor (possibly a supporting role) has.
imdb_score: The IMDb score or rating of the movie.
aspect_ratio: The aspect ratio of the movie.

movie_facebook_likes: The number of Facebook likes for the movie's Facebook page.

Lights, Camera, Data!

"Welcome to a Data-Infused Cinematic Adventure"

"Join us as we embark on an analytical voyage through the world of movies."
"Discover the Hidden Insights in IMDb Data That Define Movie Excellence."

Data Cleaning and Data Processing

Data Cleaning

Delete the blank rows and columns and delete Duplicate

Engagement

Shape Of Dataset I received with 5043,28.After cleaning 3723,28

02
Presentation

"What factors influence the success of a movie on IMDB?"

Movie Genre Analysis
Analysing and procedure

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

Find the Most Popular Genres The COUNTIF function in Excel may be used to determine the number of films in each genre.

Here is how to go about it: You may divide numerous genres in the 'genres' column into different cells by using the Text to Columns tool in a new column.

How to do it: Choose the 'genres' column.

Click the "Data" tab, then select "Text to Columns. "Select "Delimited" and then select "Next."By default, a comma is chosen as the delimiter.

Then click "Next. "Select the cell you want the split genres to go to, then click "Finish.“

Now, in a new column, you can use COUNTIF to count the occurrences of each genre. Create a list of unique genres in a separate column, and then use COUNTIF to count how many times each genre appears in the split 'genres' column.

EXAMPLE : =COUNTIF($B$2:$D$6, C2)

Calculate Descriptive Statistics:

THEN I CALCULATE THE MEAN, MEDIAN,MODE VARIENCE AND STD FOR EVERY GENRE

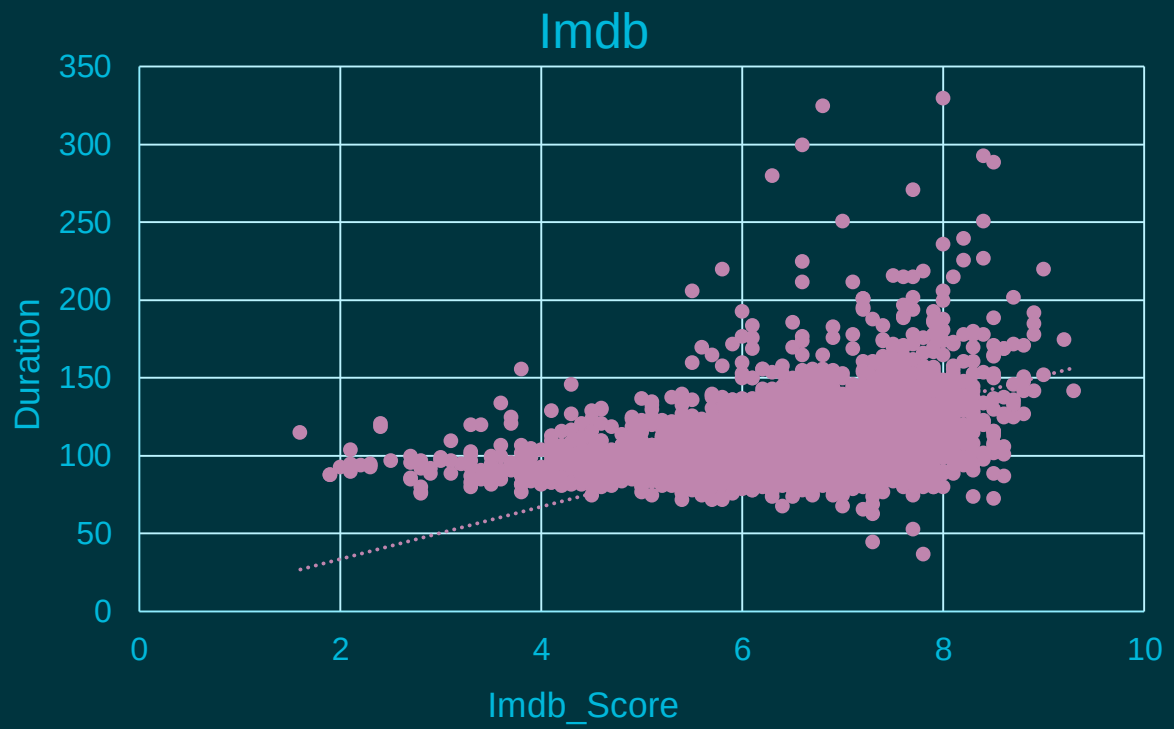
Movie Genre Analysis

Action 951	Adventure 366	Horror 159	Family 3	Biography 204
Sci-Fi 7	Drama 659	Romance 1	Thriller 1	Western 2
Musical 2	Documentary 26	Crime 253	Fantasy 37	Animation 45
Mystery 23	Comedy 984			
AVERAGE 219	MEDIAN 37	MODE 1	MAX 984	MIN 1
VAR 110070.625	STD 331.7689331			

Movie Duration Analysis Procedure

Determine Descriptive Statistics for Duration:
Choose a cell to show the mean (average) of movie running times in.
To determine the mean, enter the formula shown below:
excel Coding example: =AVERAGE(B2:B3724)
The range of movie lengths in your collection is represented by B2:B3724.
To display the median of movie lengths, choose another cell. To determine the median, enter the formula shown below:=MEDIAN(B2:B3724)in Excel The median running time of the films is determined using this formula.
Choose a different cell if you want to compute the movie duration standard deviation. Enter the formula shown below to determine the standard deviation:=STDEV(B2:B3724) in an excel copy This formula determines the movie durations' standard deviation, which assesses how widely distributed the durations are from the mean.
Scatter plot create:
Make a scatter plot comparing movie duration to IMDb rating. Columns A (duration) and B (IMDb ratings) cells should be selected. Navigate to Excel's "Insert" tab.Select "Scatter with Straight Lines" from the drop-down menu under "Scatter Chart" in the "Charts" category. The correlation between the length of the film and its IMDb rating will be displayed in a scatter plot.
Include a Trend line:
To pick the whole data series, click on one of the scatter plot's data points.
Select "Add Trend line" from the context menu by performing a right-click on the data points.

Movie Duration Analysis



Average
110.2635

Median
106

STD
22.67832

Language Analysis procedure

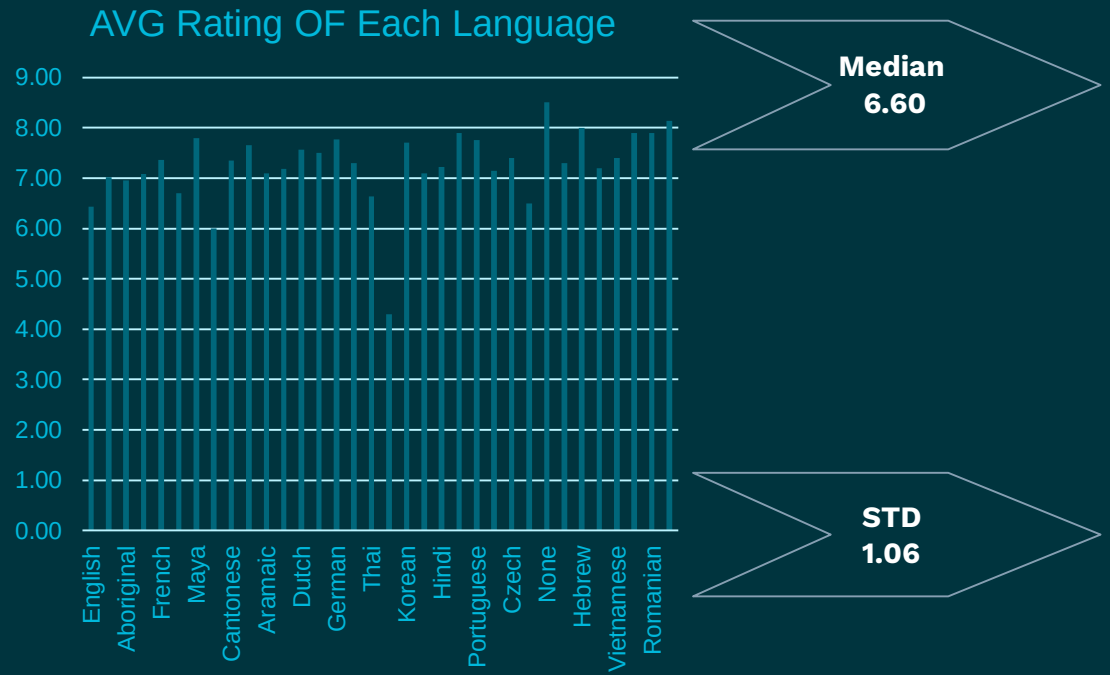
Here's how to accomplish it if you have a list of distinct languages in column A and you want to compute and display the mean IMDb ratings in column B:Make a list of Special Languages in
Step 1 List every distinct language you identified in your dataset in column I. Either manually or by utilizing Excel's "Remove Duplicates" tool, you may accomplish this.
Count of each unique language `COUNTIFS($A$2:$A$3724,D2)`

Calculate the Mean IMDb Scores for Each Language
In **Step 2** To determine the mean IMDb score for the first language (assuming "English"), insert the formula below in cell G2`=AVERAGEIFS($B$2:$B$3724,$A$2:$A$3724,D2)`in Excel The column that contains IMDb score is denoted by the symbol B:B.
The column that lists the languages is denoted by the symbol A:A.
The first distinct language in your D2 references

And for Median and STD I used `=MEDIAN($B$2:$B$3724,$A$2:$A$3724,D2)` and `=STDEV.P($B$2:$B$3724,$A$2:$A$3724,D2)` respectively

Language Analysis

Unique language	count language
English	3566
Mandarin	14
Aboriginal	2
spanish	23
French	34
Filipino	1
Maya	1
Kazakh	1
Cantonese	7
Japanese	10
Aramaic	1
Italian	7
Dutch	3
Dari	2
German	10
Mongolian	1
Thai	3
Bosnian	1
Korean	5
Hungarian	1
Hindi	5
Danish	3
Portuguese	5
Norwegian	4
Czech	1
Russian	1
None	1
Zulu	1
Hebrew	1
Arabic	1
Vietnamese	1
Indonesian	2
Romanian	1
Persian	3



Director Analysis

Calculate the Director's Average IMDb Score:
Make a new column and enter the average IMDb rating for each director in it. Assume you begin at cell C2 of column C to build this column.

Use the AVERAGEIF function in cell C2 to determine the first director's average IMDb score (assume "James Cameron" is the first director's name in column B). The equation would be as follows:

Copy the following code: `=AVERAGEIF($A$2:$A$3724,A5,$B$2:$B$3724)`The average IMDb rating for James Cameron-directed films is determined using this algorithm.

To determine the average IMDb score for each director in your dataset, copy this formula down for each director.

Utilise percentile calculations while analysing:
Enter the formula below in a cell to determine the percentile rank of a certain director's average IMDb score within the dataset:
The range of average IMDb ratings for all the directors in your dataset is represented by the code `=PERCENTRANK.INC($D$2:$D$3724,D5)`

Director Analysis

Akira Kurosawa
Avg Imdb Score – 8.70
Percentile-1.00

Tony Kaye
Avg Imdb Score – 8.60
Percentile-0.99

Alfred Hitchcock and Majid Majidi
Avg Imdb Score – 8.50
Percentile-0.998

Budget Analysis

Do the correlation coefficient calculation.

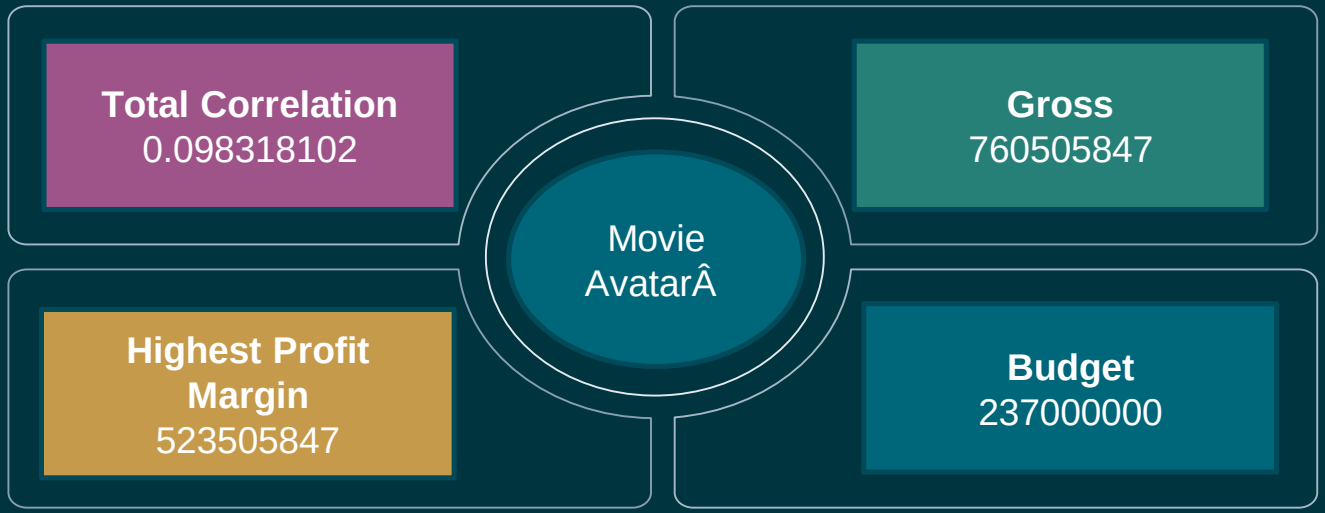
Enter the following formula in a cell to get the correlation coefficient between the Budget (column C) and the gross (column D):
Formula: =CORREL(C2:C3724,D2:D3724)

Determine the profit margin
Formula: =D2-C2

For the Highest Profit Margin I have used
Formula: =MAX(G2:G3724)

the movie with the highest profit margin
Formula: =INDEX(A2:A3724,MATCH(H2,G2:G3724,0))

Budget Analysis



«Believe in yourself and all that you are. Know that there is something inside you that is greater than any obstacle.»

—Christian D. Larson

Thanks!

Do you have any questions?

My_official2o23@outlook.com
91 6295138110

You can find the Working file

https://docs.google.com/spreadsheets/d/1VkAWCzAdYm9GTSy-oUzykZt-PCKMX28N/edit?usp=drive_link&ouid=113119245928557471712&rtpof=true&sd=true