

[Click Here for the VIDEO](#)



PROJECT 5

IMDB MOVIE ANALYSIS

SUBMITTED BY

SWATHY MUGUNDAN

Problem Statement

- 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.
- The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.
- After analysis, a report to be created that tells a story with data.
- This will include initial problem, findings, and the insights gained.
- Data Visualizations to be used to make findings more understandable.

Project Goal



To use the knowledge of statistics and Excel to provide insights that can drive a decision-making.

The entire analysis aims to provide actionable insights that can help the stakeholders make informed decisions.

Tech Stack Used



MS - EXCEL

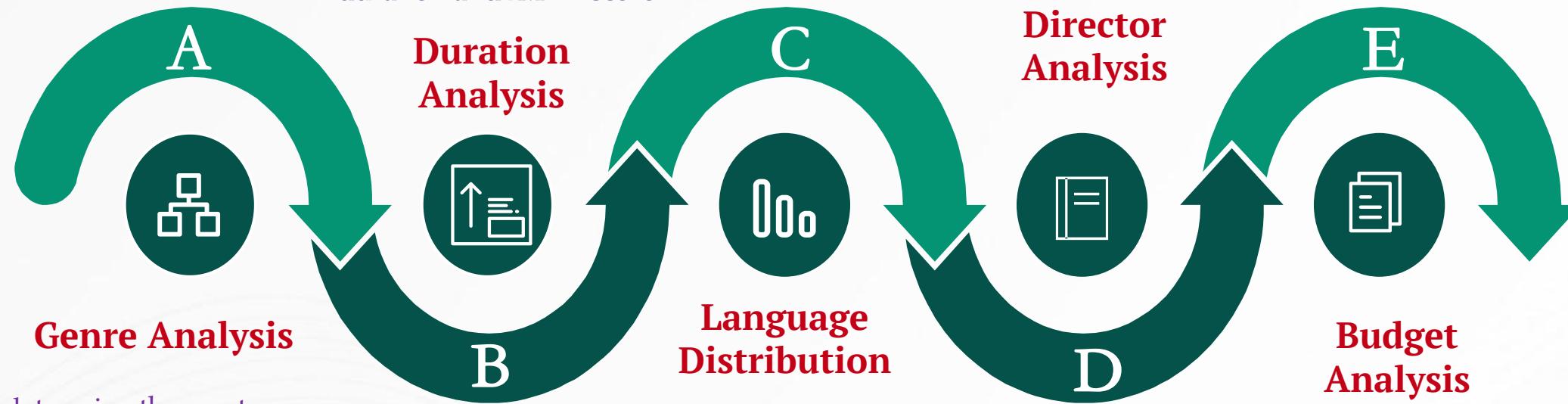
- MS Excel is a spreadsheet program where one can record data in the form of tables.
- It is easy to analyse data in an Excel spreadsheet.
- A spreadsheet is in the form of a table comprising rows and columns.

Benefits of MS - EXCEL

- ✓ Easy to store and manage data
- ✓ Application of Mathematical and Statistical formulas
- ✓ More Secured
- ✓ Clearer visibility of information
- ✓ Data Processing Application

Data Analytics Tasks

[Click Here for the VIDEO](#)



To determine the most common genres of movies in the dataset and calculate descriptive statistics based on their IMDB Scores

To analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

Language Distribution

To identify the top directors based on their average IMDB score and analyze their contribution to the success of movies

Budget Analysis

To analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Descriptive Statistics

- Descriptive statistics refers to a branch of statistics that involves summarizing, organizing, and presenting data **meaningfully and concisely**.
- The primary goal of descriptive statistics is to provide a clear and concise summary of the data, enabling researchers or analysts to gain insights and **understand patterns, trends, and distributions within the dataset**.
- This summary typically includes measures such as **central tendency** (e.g., mean, median, mode), **dispersion** (e.g., range, variance, standard deviation), and **shape of the distribution** (e.g., skewness, kurtosis).
- Descriptive statistics also involves a graphical representation of data through **charts, graphs, and tables**, which can further aid in visualizing and interpreting the information.

Data Cleaning Strategy



- This step involves preprocessing the data to make it suitable for analysis.
- It includes handling missing values, removing duplicates, converting data types if necessary, and possibly feature engineering.

Why Data Cleaning is Necessary ?

It involves identifying and removing any missing, duplicate, or irrelevant data. This ultimately increase overall productivity and allow for the highest quality information in decision-making.

It also helps in removal of errors when multiple sources of data are at play.

Data Cleaning : Removal of Duplicate Data

Select the entire data from the address bar. Select Data -> Remove Duplicates.

Select “All Columns” in the dialogue box and click “OK”. Duplicate data are removed from the dataset.

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Movie_Title	Director_Name	Duration	Genres	Language	imdb_score	Budget	Gross									
2	Avatar	James Cameron	178	Action Adventu	English	7.9	237000000	760505847									
3	Pirates of the Caribbean: At World's End	Gore Verbinski	169	Action Adventu	English	7.1	300000000	309404152									
4	Spectre	Sam Mendes	148														
5	The Dark Knight Rises	Christopher Nolan	164														
6	Star Wars: Episode VII - The Force Awakens	Doug Walker															
7	John Carter	Andrew Stanton	132														
8	Spider-Man 3	Sam Raimi	156														
9	Tangled	Nathan Greno	100														
10	Avengers: Age of Ultron	Joss Whedon	141														
11	Harry Potter and the Half-Blood Prince	David Yates	153														
12	Batman v Superman: Dawn of Justice	Zack Snyder	183	Action Adventu	English	6.3	250000000	330245002									
13	Superman Returns	Bryan Singer	169	Action Adventu	English	6.1	209000000	200069408									
14	Quantum of Solace	Marc Forster	106	Action Adventu	English	6.7	200000000	168368427									
15	Pirates of the Caribbean: Dead Man's Chest	Gore Verbinski	151	Action Adventu	English	7.3	225000000	423032628									
16	The Lone Ranger	Gore Verbinski	150	Action Adventu	English	6.5	215000000	89289910									
17	Man of Steel	Zack Snyder	143	Action Adventu	English	7.2	225000000	291021565									
18	The Chronicles of Narnia: Prince Caspian	Andrew Adamson	150	Action Adventu	English	6.6	225000000	141614023									
19	The Avengers	Joss Whedon	173	Action Adventu	English	8.1	220000000	623279547									
20	Pirates of the Caribbean: On Stranger Tides	Rob Marshall	136	Action Adventu	English	6.7	250000000	241063875									
21	Men in Black 3	Barry Sonnenfeld	106	Action Adventu	English	6.8	225000000	179020854									
22	The Hobbit: The Desolation of Smaug	Peter Jackson	164	Adventure Fan	English	7.5	250000000	255108370									
23	The Amazing Spider-Man	Marc Webb	153	Action Adventu	English	7	230000000	262030663									

A tooltip window titled "Excel" is displayed, stating: "122 duplicate values found and removed; 4921 unique values remain." It includes a link "Give Feedback to Microsoft" and an "OK" button.

Data Cleaning : Removing Unnecessary Columns

Try to figure out the unwanted attributes from the dataset and remove the columns completely.

This helps us to focus on the important attributes for data analysis and will increase the accuracy of decision-making.

Identified Unnecessary Attributes : Color, num_critic_for_rev, director_facebook_likes, actor_3_facebook_likes, actor_1_facebook_likes, num_voted_users, cast_total_facebook_likes, facenumber_in_poster, plot_keywords, movie_imdb_link, num_user_for_reviews, content_rating, actor_2_facebook_likes, aspect_ratio, movie_facebook_likes, actor_1_name, actor_2_name, actor_3_name, country

Data Cleaning : Removing Blank Cells

Select the entire data and click on Data -> Filter. For each field, identify the blank cells and try to give default values based on the data type of the attributes.

Setting up default values

director_name : blank to NULL

duration : blank to zero (0)

gross, budget : blank to zero (0)

genre : No blank cells

imdb_score, movie_title : No blank cells

language: blank to NULL

	A	B	C	D	E	F	G	H
1	movie_title	director_name	duration	genres	language	imdb_score	budget	gross
2	Avatar	James Cameron	178	Action Adventure Fantasy Sci-Fi	English	7.9	237000000	760505847
3	Pirates of the Caribbean: At World's End	Gore Verbinski	169	Action Adventure Fantasy	English	7.1	300000000	309404152
4	Spectre	Sam Mendes	148	Action Adventure Thriller	English	6.8	245000000	200074175
5	The Dark Knight Rises	Christopher Nolan	164	Action Thriller	English	8.5	250000000	448130642
6	Star Wars: Episode VII - The Force Awakens	Doug Walker	0	Documentary	NULL	7.1	0	0
7	John Carter	Andrew Stanton	132	Action Adventure Sci-Fi	English	6.6	263700000	73058679
8	Spider-Man 3	Sam Raimi	156	Action Adventure Romance	English	6.2	258000000	336530303
9	Tangled	Nathan Greno	100	Adventure Animation Comedy Family	English	7.8	260000000	200807262
10	Avengers: Age of Ultron	Joss Whedon	141	Action Adventure Sci-Fi	English	7.5	250000000	458991599
11	Harry Potter and the Half-Blood Prince	David Yates	153	Adventure Family Fantasy Mystery	English	7.5	250000000	301956980
12	Batman v Superman: Dawn of Justice	Zack Snyder	183	Action Adventure Sci-Fi	English	6.9	250000000	330249062
13	Superman Returns	Bryan Singer	169	Action Adventure Sci-Fi	English	6.1	209000000	200069408
14	Quantum of Solace	Marc Forster	106	Action Adventure	English	6.7	200000000	168368427
15	Pirates of the Caribbean: Dead Man's Chest	Gore Verbinski	151	Action Adventure Fantasy	English	7.3	225000000	423032628
16	The Lone Ranger	Gore Verbinski	150	Action Adventure Western	English	6.5	215000000	89289910
17	Man of Steel	Zack Snyder	143	Action Adventure Fantasy Sci-Fi	English	7.2	225000000	291021565
18	The Chronicles of Narnia: Prince Caspian	Andrew Adamson	150	Action Adventure Family Fantasy	English	6.6	225000000	141614023
19	The Avengers	Joss Whedon	173	Action Adventure Sci-Fi	English	8.1	220000000	623279547
20	Pirates of the Caribbean: On Stranger Tides	Rob Marshall	136	Action Adventure Fantasy	English	6.7	250000000	241063875
21	Men in Black 3	Barry Sonnenfeld	106	Action Adventure Comedy Family Fai	English	6.8	225000000	179020854
22	The Hobbit: The Desolation of Smaug	Peter Jackson	164	Action Adventure Fantasy	English	7.5	250000000	255108370
23	The Amazing Spider-Man	Marc Webb	153	Action Adventure Fantasy	English	7	230000000	262030663

Sample Data after Data Cleaning

Task 1 : Movie Genre Analysis

- **Objective :** To 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.
- **Approach :**
 - ✓ Use text-split function to split the individual genre of a movie
 - ✓ Using countif, count the number of movies under each genre
 - ✓ Calculate the descriptive statistics (using filter and formulas) for each genre
 - ✓ Compare the statistics with ratings and observe the impact

Task 1 : Movie Genre Analysis

Genre	Count	Mean	MAX	MIN	Range	Median	Variance	SD	Overall Profit
Action	1113	6.231626235	9.1	1.7	7.4	6.6	1.271906	1.127788	3631406647
Adventure	889	6.436332958	8.9	1.9	7	6.6	1.263822	1.124199	8427389733
Animation	240	6.575	8.6	1.7	6.9	6.6	1.26431	1.124416	-83472470
Biography	291	7.148797251	8.9	4.5	4.4	6.6	1.272219	1.127927	1758159695
Comedy	1848	6.192857143	9.5	1.7	7.8	6.6	1.271581	1.127644	12483835284
Crime	869	6.56306099	9.3	2.4	6.9	6.6	1.276305	1.129737	131660874
Documentary	120	7.180833333	8.7	1.6	7.1	6.6	1.270658	1.127235	647496972
Drama	2537	6.765037446	9.3	2	7.3	6.6	1.27337	1.128437	-8192218342
Family	533	6.234709193	8.7	1.7	7	6.6	1.269831	1.126868	9394670766
Fantasy	580	6.296034483	8.9	1.7	7.2	6.6	1.270105	1.12699	6990955491
Film-Noir	6	7.633333333	8.2	7.1	1.1	6.6	1.829041	1.35242	-7776450
Game-Show	1	2.9	2.9	2.9	0	2.9	0	0	-3000000
History	204	7.080882353	8.9	2	6.9	6.6	1.265173	1.124799	-2176960734
Horror	540	5.80462963	8.7	2.2	6.5	6.6	1.276188	1.129685	-7695778131
Music	322	6.448447205	8.5	1.6	6.9	6.6	1.273741	1.128601	4073740365
Musical	130	6.490769231	8.5	2.1	6.4	6.6	1.264512	1.124505	1553568714
Mystery	484	6.481818182	8.6	2.2	6.4	6.6	1.272	1.12783	3679424241
News	3	7.533333333	8.1	7.1	1	6.6	1.418632	1.191064	11946793
Reality-TV	2	4.75	6.6	2.9	3.7	6.4	1.589605	1.260795	-3000000
Romance	1082	6.445656192	8.6	2.1	6.5	6.6	1.271333	1.127534	7671227056
Sci-Fi	587	6.268143101	8.8	1.9	6.9	6.6	1.272376	1.127997	-9186051170
Short	5	6.38	7.1	5.2	1.9	6.6	1.412833	1.188626	1805534
Sport	177	6.601129944	8.7	2	6.7	6.6	1.267398	1.125788	1824945448
Thriller	1362	6.305580029	9	2.2	6.8	6.6	1.271541	1.127626	3823161521
War	210	7.070952381	8.6	2.7	5.9	6.6	1.268413	1.126239	-4121940918
Western	93	6.705376344	8.9	3.8	5.1	6.6	1.271783	1.127734	126878955

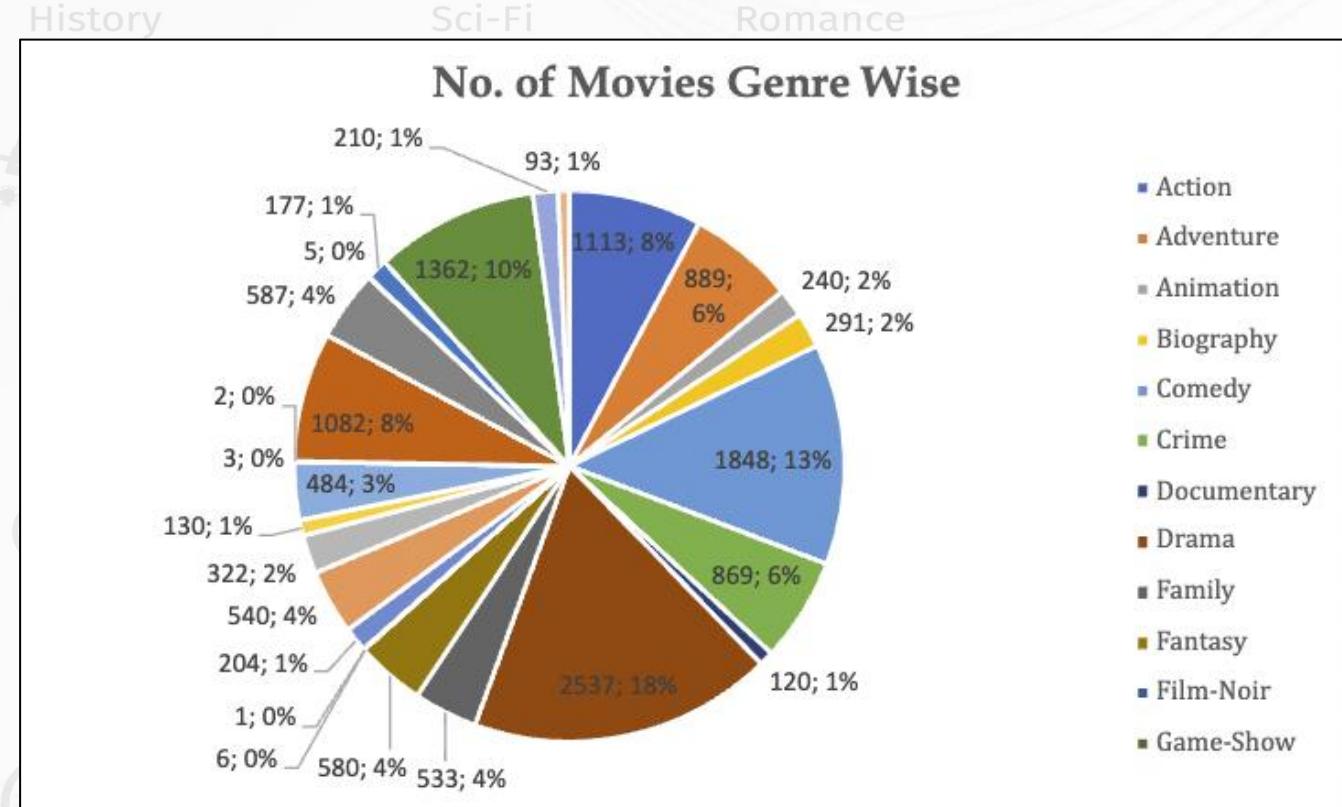
Descriptive Statistics

Average/Mean	547.2307692
Median	306.5
Variance	390777.6246
Std. Deviation	625.1220878
Maximum	2537
Minimum	1
Range	2536

[CLICK HERE TO VIEW THE EXCEL SHEET](#)

Task 1 : Movie Genre Analysis

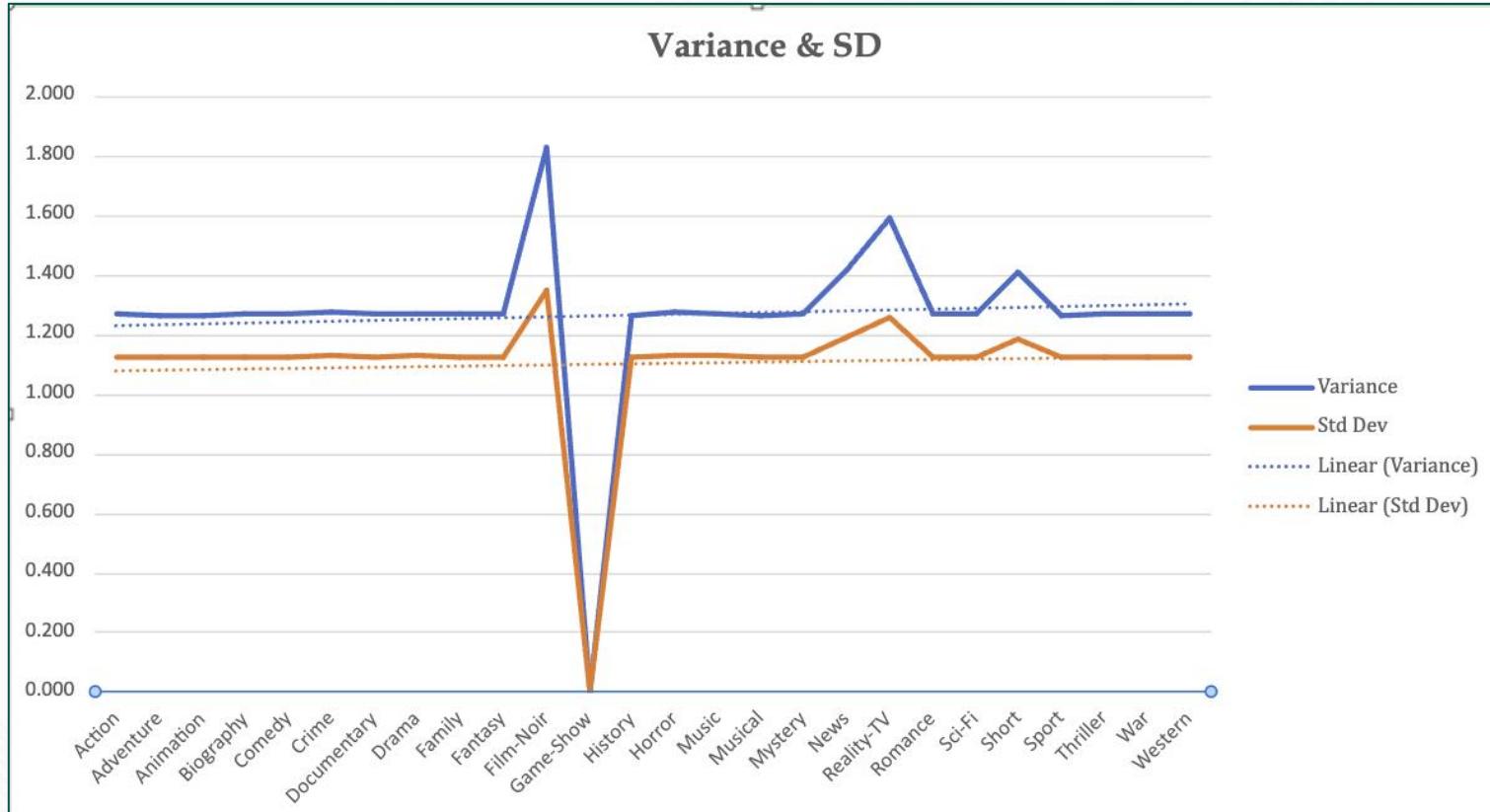
Genre	No. of Movies
Action	1113
Adventure	889
Animation	240
Biography	291
Comedy	1848
Crime	869
Documentary	120
Drama	2537
Family	533
Fantasy	580
Film-Noir	6
Game-Show	1
History	204
Horror	540
Music	322
Musical	130
Mystery	484
News	3
Reality-TV	2
Romance	1082
Sci-Fi	587
Short	5
Sport	177
Thriller	1362
War	210
Western	93



[CLICK HERE TO VIEW THE EXCEL SHEET](#)

Task 1 : Movie Genre Analysis

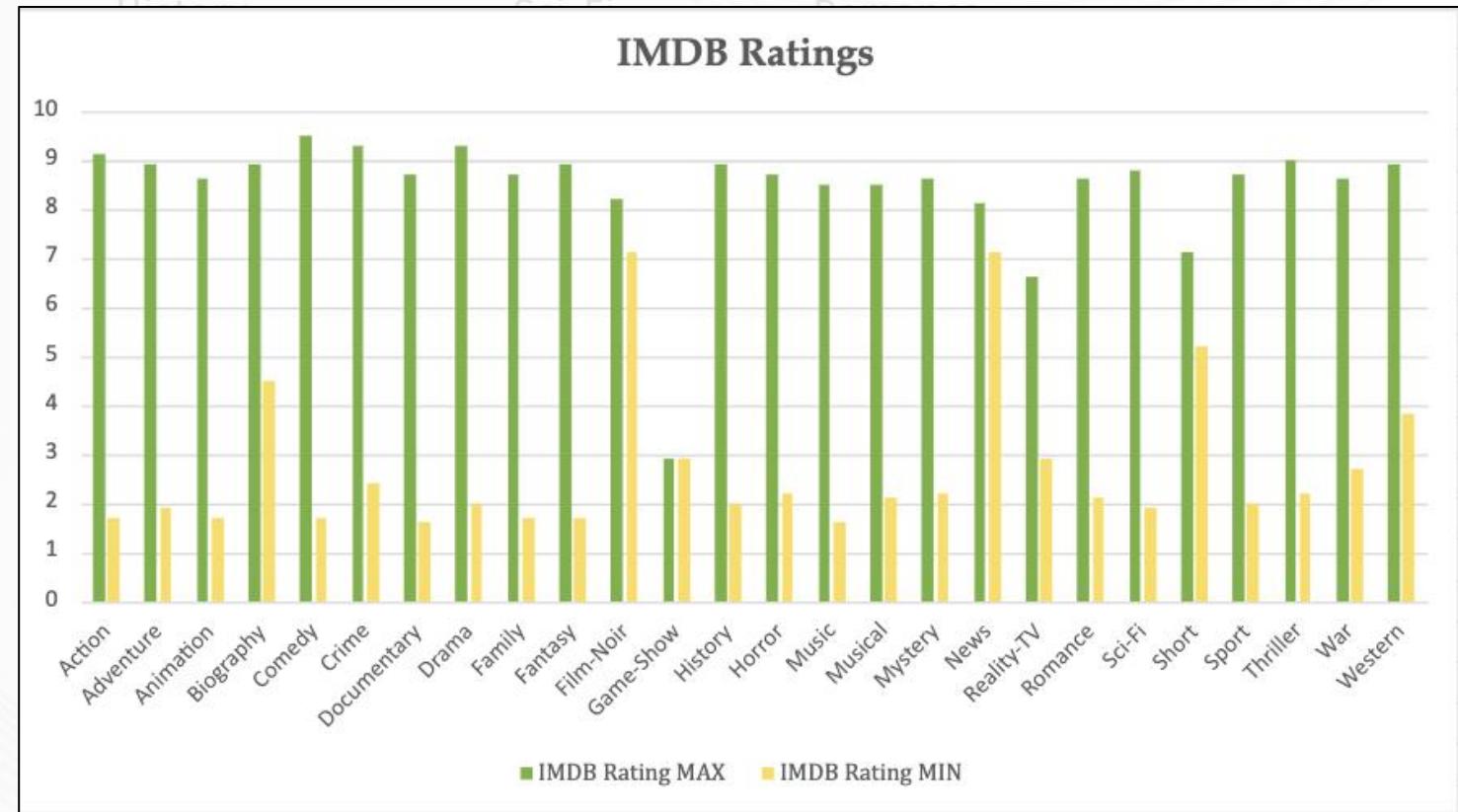
Genre	IMDB RATING	
	Variance	SD
Action	1.272	1.128
Adventure	1.264	1.124
Animation	1.264	1.124
Biography	1.272	1.128
Comedy	1.272	1.128
Crime	1.276	1.130
Documentary	1.271	1.127
Drama	1.273	1.128
Family	1.270	1.127
Fantasy	1.270	1.127
Film-Noir	1.829	1.352
Game-Show	0.000	0.000
History	1.265	1.125
Horror	1.276	1.130
Music	1.274	1.129
Musical	1.265	1.125
Mystery	1.272	1.128
News	1.419	1.191
Reality-TV	1.590	1.261
Romance	1.271	1.128
Sci-Fi	1.272	1.128
Short	1.413	1.189
Sport	1.267	1.126
Thriller	1.272	1.128
War	1.268	1.126
Western	1.272	1.128



[CLICK HERE TO VIEW THE EXCEL SHEET](#)

Task 1 : Movie Genre Analysis

Genre	IMDB Rating	
	MAX	MIN
Action	9.1	1.7
Adventure	8.9	1.9
Animation	8.6	1.7
Biography	8.9	4.5
Comedy	9.5	1.7
Crime	9.3	2.4
Documentary	8.7	1.6
Drama	9.3	2
Family	8.7	1.7
Fantasy	8.9	1.7
Film-Noir	8.2	7.1
Game-Show	2.9	2.9
History	8.9	2
Horror	8.7	2.2
Music	8.5	1.6
Musical	8.5	2.1
Mystery	8.6	2.2
News	8.1	7.1
Reality-TV	6.6	2.9
Romance	8.6	2.1
Sci-Fi	8.8	1.9
Short	7.1	5.2
Sport	8.7	2
Thriller	9	2.2
War	8.6	2.7
Western	8.9	3.8



[CLICK HERE TO VIEW THE EXCEL SHEET](#)

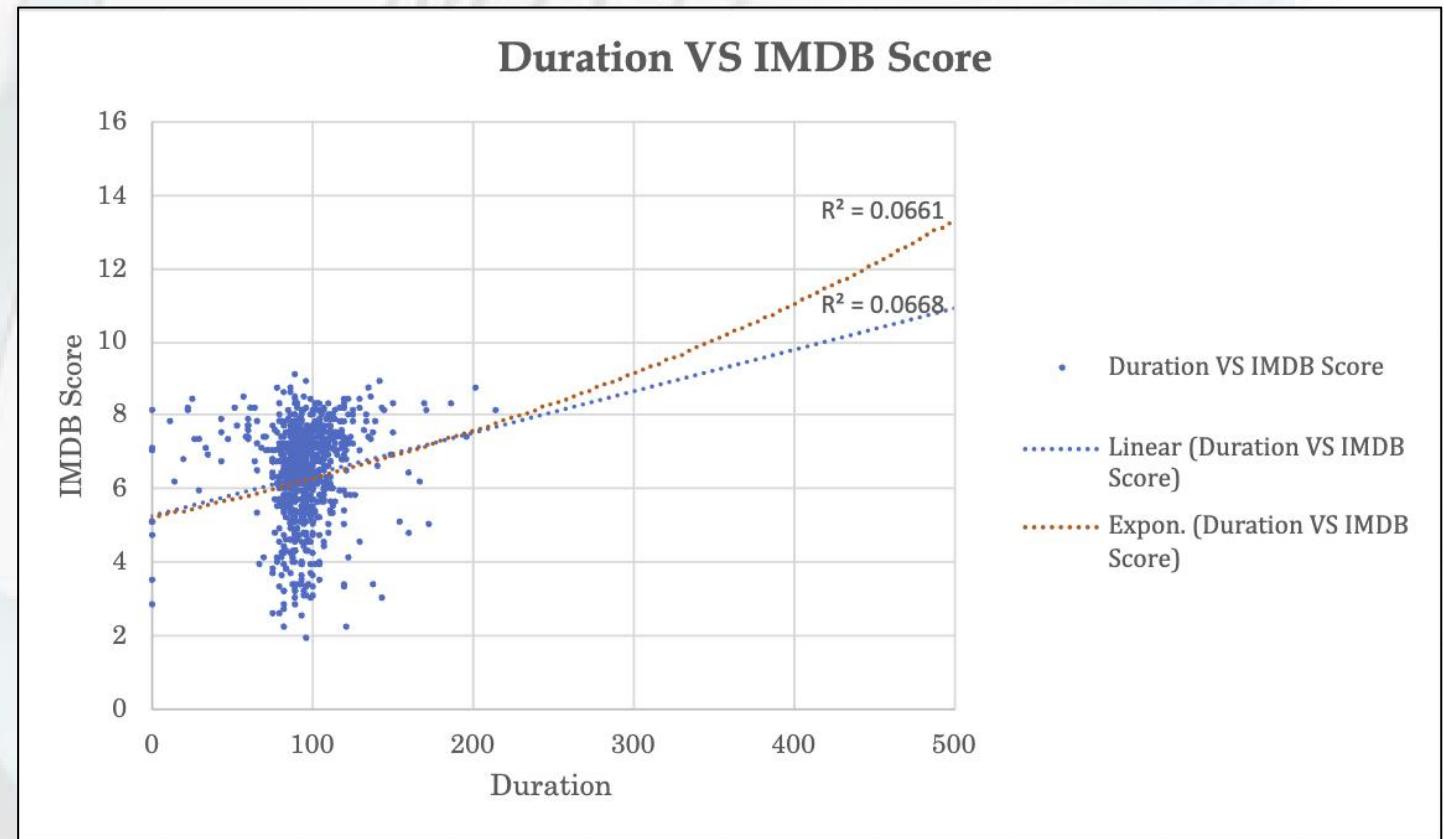
Task 2 : Movie Duration Analysis

- **Objective :** To analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score
- **Approach :**
 - ✓ Calculation of descriptive statistics for the movie duration column
 - ✓ Creating a scatter plot to visualise the relationship between the movie durations and the IMDB Score of the movies
 - ✓ Observing the trend of the data spread using trendline

Task 2 : Movie Duration Analysis

MOVIE DURATION ANALYSIS	
Metric	Values
Average	106.7764682
Median	103
Mode	90
Std Dev	25.92551132
Variance	672.1321372
Maximum	511
Minimum	0

R^2 is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable.



[CLICK HERE TO VIEW THE EXCEL SHEET](#)

Task 3 : Language Analysis

- **Objective :** To determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.
- **Approach :**
 - ✓ Using the Pivot Table, the number of movies which belong to a particular language is calculated
 - ✓ The descriptive statistics is calculated for IMDB Scores for each language
 - ✓ Observing the results to understand the relationship between language and movie ratings

Task 3 : Language Analysis

PivotTable Fields X

Choose fields to add to the report and drag them between the areas below:

⚙️ 🔍 Search

Rows

movie_title
 director_name
 duration
 genres
 language
 imdb_score
 budget
 gross

Columns

Values

Σ Values

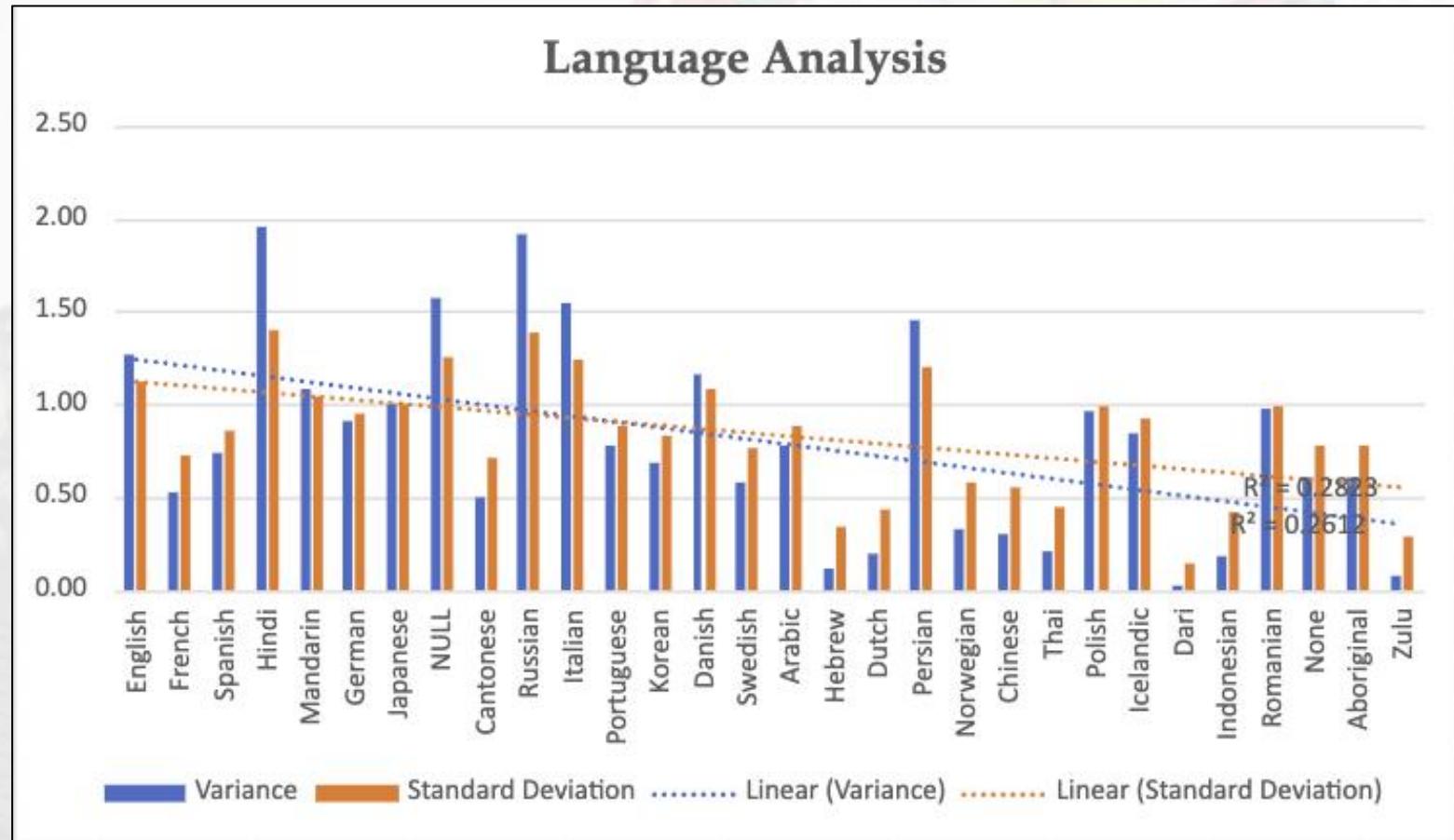
Count
Average
Max
Min
Var
SD

language	Count	Average	Max	Min	Var	SD
English	4586	6.39	9.50	1.60	1.27	1.13
French	73	7.04	8.40	4.90	0.53	0.73
Spanish	40	6.94	8.20	4.40	0.73	0.86
Hindi	28	6.63	8.50	2.80	1.96	1.40
Mandarin	24	6.79	7.90	3.20	1.08	1.04
German	19	7.34	8.50	4.90	0.91	0.95
Japanese	17	7.35	8.70	5.60	1.00	1.00
NULL	12	6.85	9.10	4.80	1.57	1.25
Cantonese	11	6.95	7.80	5.30	0.50	0.70
Russian	11	6.36	8.10	4.10	1.91	1.38
Italian	11	7.23	8.90	5.10	1.55	1.24
Portuguese	8	7.49	8.70	6.10	0.78	0.88
Korean	8	7.39	8.40	5.70	0.68	0.83
Danish	5	7.50	8.30	5.70	1.16	1.08
Swedish	5	7.44	8.20	6.60	0.57	0.76
Arabic	5	7.38	8.20	6.00	0.78	0.88
Hebrew	5	7.58	8.00	7.20	0.11	0.33
Dutch	4	7.43	7.80	7.00	0.19	0.43
Persian	4	7.58	8.50	5.90	1.45	1.20
Norwegian	4	7.15	7.60	6.40	0.33	0.57
Chinese	3	5.67	6.20	5.10	0.30	0.55
Thai	3	6.63	7.10	6.20	0.20	0.45
Polish	3	7.97	9.10	7.40	0.96	0.98

[CLICK HERE TO VIEW THE EXCEL SHEET](#)

Task 3 : Language Analysis

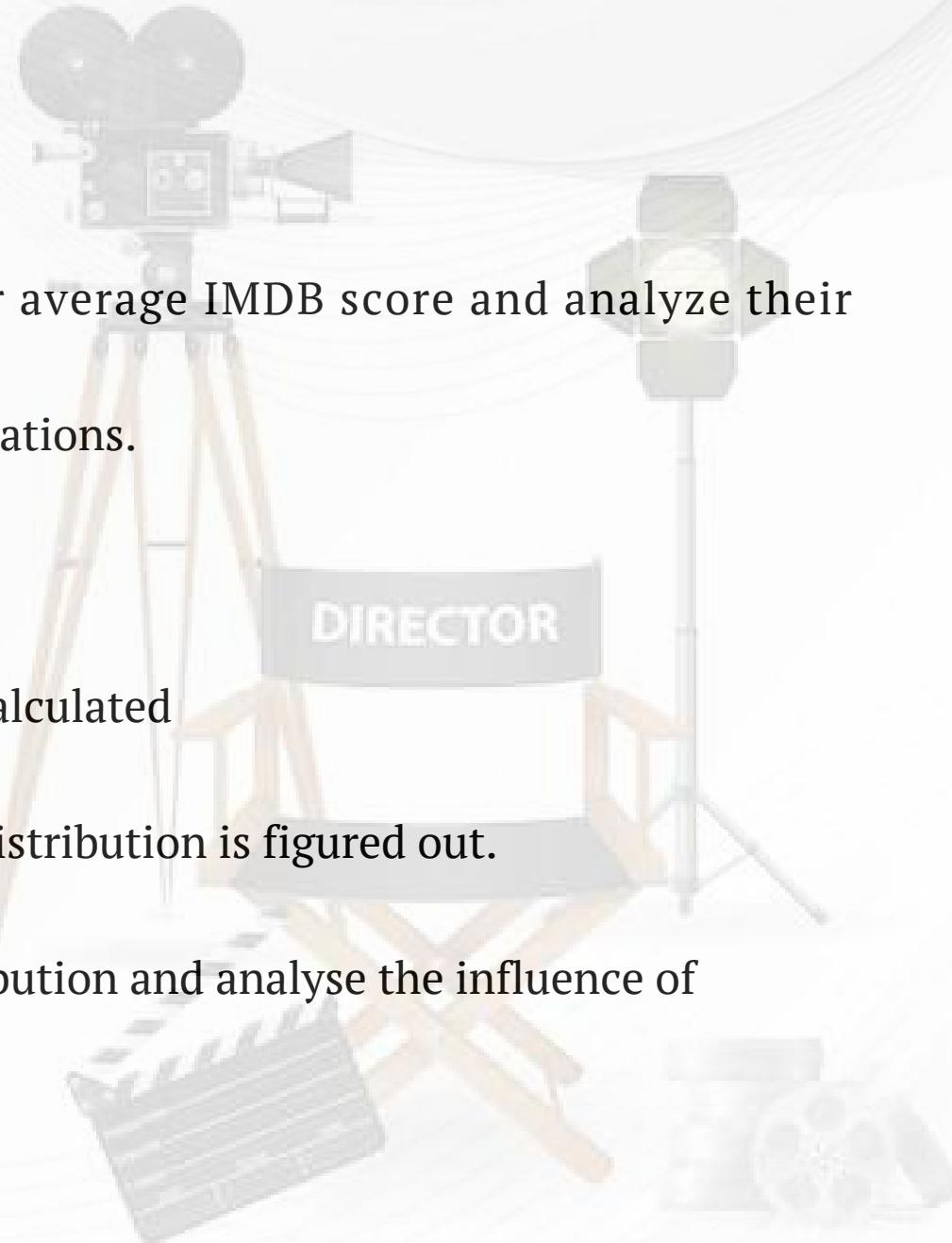
Language	Var	SD
English	1.27	1.13
French	0.53	0.73
Spanish	0.73	0.86
Hindi	1.96	1.40
Mandarin	1.08	1.04
German	0.91	0.95
Japanese	1.00	1.00
NULL	1.57	1.25
Cantonese	0.50	0.70
Russian	1.91	1.38
Italian	1.55	1.24
Portuguese	0.78	0.88
Korean	0.68	0.83
Danish	1.16	1.08
Swedish	0.57	0.76
Arabic	0.78	0.88
Hebrew	0.11	0.33
Dutch	0.19	0.43
Persian	1.45	1.20
Norwegian	0.33	0.57
Chinese	0.30	0.55
Thai	0.20	0.45
Polish	0.96	0.98
Icelandic	0.84	0.92
Dari	0.02	0.14
Indonesian	0.18	0.42
Romanian	0.98	0.99
None	0.61	0.78
Aboriginal	0.61	0.78
Zulu	0.08	0.28



[CLICK HERE TO VIEW THE EXCEL SHEET](#)

Task 4 : Director Analysis

- **Objective :** To identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.
- **Approach :**
 - ✓ Using the Pivot Table, the average of IMDB Scores are calculated
 - ✓ Using the % Grand Total, the contribution to the total distribution is figured out.
 - ✓ Compare the scores of the directors to the overall distribution and analyse the influence of directors on movie ratings.



Task 4 : Director Analysis

PivotTable Fields

Choose fields to add to the report and drag them between the areas below:

Search Filters

director_name imdb_score

Rows: director_name

Columns: Values

Values: Average, % Grand Total, Count

director_name	Average	% Grand Total	Count
John Blanchard	9.5	109.45%	1
Sadyk Sher-Niyaz	8.7	100.23%	1
Mitchell Altieri	8.7	100.23%	1
Cary Bell	8.7	100.23%	1
Mike Mayhall	8.6	99.08%	1
Charles Chaplin	8.6	99.08%	1
Damien Chazelle	8.5	97.93%	1
Ron Fricke	8.5	97.93%	1
Raja Menon	8.5	97.93%	1
Majid Majidi	8.5	97.93%	1
Grand Total	8.68	100.00%	10

Top 10 Directors based on
their Average IMDB Ratings

director_name	Average	% Grand Total	Count
Martin Scorsese	7.66	106.37%	20
NULL	7.5333333	104.61%	102
Steven Spielberg	7.4807692	103.88%	26
Clint Eastwood	7.225	100.33%	20
Ridley Scott	7.13125	99.03%	16
Tim Burton	7.05	97.90%	14
Woody Allen	7.0090909	97.33%	22
Oliver Stone	6.95	96.51%	14
Steven Soderbergh	6.68	92.76%	15
Spike Lee	6.56875	91.21%	16
Renny Harlin	5.7466667	79.80%	15
Grand Total	7.2014286	100.00%	280

Top 10 Directors based on
their Movies Count

[CLICK HERE TO VIEW THE EXCEL SHEET](#)

Task 5 : Budget Analysis

- **Objective :** To analyze the correlation between movie budgets and gross earnings, and to identify the movies with the highest profit margin.
- **Approach :**
 - ✓ Using the correl formula, to calculate the correlation coefficient between the movie budget and the gross earnings.
 - ✓ Calculate the profit of each movie by finding the difference between the budget and the gross
 - ✓ Find the movie with highest profit margin using the max function

Task 5 : Budget Analysis

movie_title	budget	gross	profit
Avatar	237000000	760505847	523505847
Pirates of the Caribbean: At World's End	300000000	309404152	9404152
Spectre	245000000	200074175	-44925825
The Dark Knight Rises	250000000	448130642	198130642
Star Wars: Episode VII - The Force Awakens	0	0	0
John Carter	263700000	73058679	-190641321
Spider-Man 3	258000000	336530303	78530303
Tangled	260000000	200807262	-59192738
Avengers: Age of Ultron	250000000	458991599	208991599
Harry Potter and the Half-Blood Prince	250000000	301956980	51956980
Batman v Superman: Dawn of Justice	250000000	330249062	80249062
Superman Returns	209000000	200069408	-8930592
Quantum of Solace	200000000	168368427	-31631573
Pirates of the Caribbean: Dead Man's Chest	225000000	423032628	198032628
The Lone Ranger	215000000	89289910	-125710090
Man of Steel	225000000	291021565	66021565
The Chronicles of Narnia: Prince Caspian	225000000	141614023	-83385977
The Avengers	220000000	623279547	403279547
Pirates of the Caribbean: On Stranger Tides	250000000	241063875	-8936125
Men in Black 3	225000000	179020854	-45979146
The Hobbit: The Desolation of Smaug	250000000	255108370	5108370
The Amazing Spider-Man	230000000	262030663	32030663

Profit is calculated by subtracting the budget from gross values.

$$\text{Profit} = \text{Gross} - \text{Budget}$$

Hit : If the value is positive (greater than zero), then the movie is a hit.

Flop : If the value is negative (lesser than zero), then the movie is flop.

Zero : If there is no data given for budget or gross then, the profit is made as null using Zero.

[CLICK HERE TO VIEW THE EXCEL SHEET](#)

Task 5 : Budget Analysis

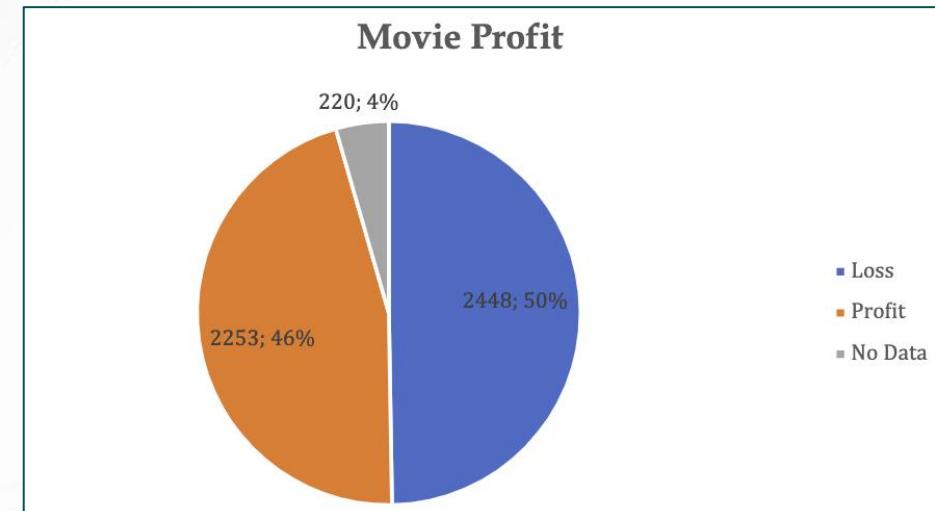
CORREL FUNCTION : The CORREL function returns the correlation coefficient of two cell ranges. Use the correlation coefficient to determine the relationship between two properties. Here, the relationship between the budget and gross attributes are figured out. **Formula : =CORREL(B2:B4922;C2:C4922)**

Observation : Correl Values between -0.3 and 0.3 is a sign of weak or no correlation.

MAX FUNCTION : The MAX function is a premade function in Excel, which finds the highest number in a range. Here, we have found out the maximum value is **523505847**

Correlation
0.117690714
Highest Profit
523505847

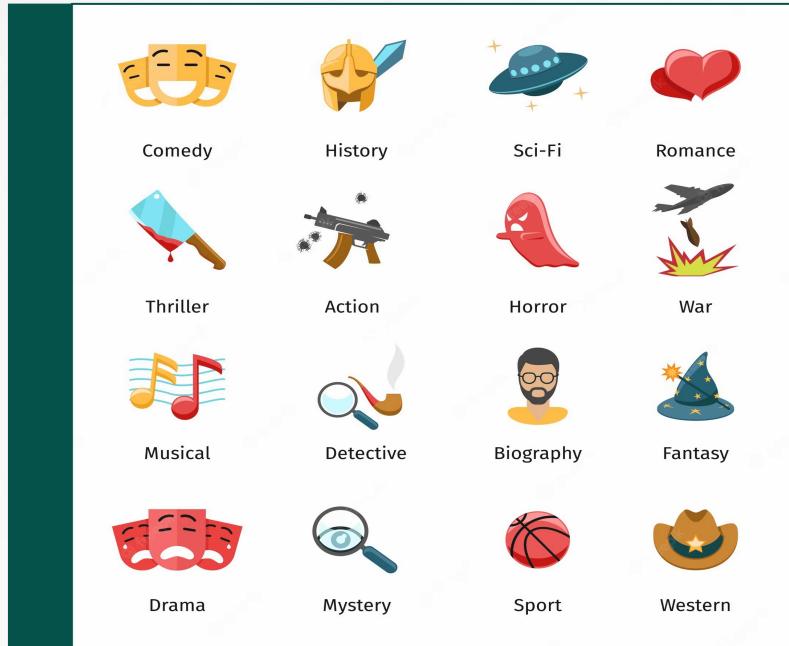
Movies	Count
Loss	2448
Profit	2253
No Data	220
Total	4921



[CLICK HERE TO VIEW THE EXCEL SHEET](#)

RESULTS & CONCLUSIONS





Genre Analysis

It is observed that, genre has an impact in the IMDB Ratings. Although most of the genre has a maximum rating above 8, based on the number of movies released on a particular genre, the maximum and minimum scores are impacted.



Duration Analysis

Based on the R2 value generated on the trendline, it is observed that, duration has 6% impact upon the ratings. Based on the scatter plot, its visualised that the data points does not fall on the regression line completely and it depicts that the variance is small.



Language Analysis

It is observed that, language has a very big impact on the IMDB Ratings. The graph plotted for Variance and SD shows that, more the number of movies, there is a good fit for data based on the ratings. Also, variance greater than 1 depicts that the values are very far from mean.



Budget Analysis

Based on the CORREL Function, it is observed that there is a weak relationship between the gross and budget attributes. The list of movies are categorised as hit or flop based on the profit values. Using Max function, the highest profit margin was found to be Avatar.



Director Analysis

- The top 10 directors based on their average IMDB Scores are figured out using the pivot table.
- But, this does not give a good influence on the ratings.
- The directors with single movie and having high rating are being the top-most which does not behave as a good way of predictions.
- Hence, one more pivot table was added to find out the top 10 directors based on the average IMDB Scores and the number of movies directed.
- The correlation between the number of movies and the average scores are calculated as **0.125 (~0.13)**
- This states that, there is a weak correlation between the director and the average ratings.
- It can be concluded that, there is **no much influence** on the ratings based on the director.

[Click Here for the VIDEO](#)

04

NAME
THE SONG

