MODULE 5

IMDB MOVIE ANALYSIS

REPORT

Gunisha Chopra
DATA ANALYST

IMDB MOVIE ANALYSIS

FINAL PROJECT-1

PROJECT DESCRIPTION

For the final project. I am provided with a dataset with various columns of different IMDB movies. I am required to Frame the problem. For this task, I need to define a problem on which I want to shed some light.

We can do this by asking 'What?' This is where I frame the problem i.e. What is the problem?

These questions guide our thinking: What do you see happening?

- What is your hypothesis for the cause of the problem? (This will be broadly based on intuition initially)
- What is the impact of the problem on stakeholders?
- What is the impact of the problem not being solved?

Answering these questions will help define a problem we are trying to solve and allow us to find the right data for further analysis.

Once we have defined a problem, we clean the data as necessary and use our Data Analysis skills to explore the data set and derive insights.

We make sure to use the 5 Whys Analysis in our analysis and use this to create a report that conveys a data story.

Once we have framed the problem and gathered initial insights from the data, we can ask the following questions as we dig deeper into our analysis.

- What do we see happening?
- What are the specific symptoms of the problem?
- What is our hypothesis for the cause of the problem?

THE FIVE 'WHYs' APPROACH

Once we have the problem better defined, we can use the 5 Whys technique to determine its root cause by repeatedly asking "Why".

It is also known as the Root Cause Analysis, developed by Sakichi Toyoda, founder of Toyota Industries. Here's an example of how this technique could be used to figure out the cause of the following problem: A business went over budget on a recent project.

Q: "Why did we go over budget on our project?"

A: It took much longer than we expected to complete.

Q: "Why did it take longer than expected to complete?"

A: We had to redesign several elements of the product.

Q: "Why did we have to redesign elements of the product?"

A: The features of the product were confusing to use.

Q: “Why were the features of the product confusing to use?”

A: We made incorrect assumptions about what users wanted.

Q: “Why did we make incorrect assumptions about what users wanted?”

A: Our user experience research team did not ask effective questions.

As we see above, what looked like a budgeting problem turned out to be a problem with the user experience team not working effectively.

While asking Why is easy, what we are interested in is the answer. Each time we answer why, the next time gets more difficult as we must think deeply about the reasons for this. As we ask why, we may find that you have multiple answers for the same question.

TECH-STACK USED:

Microsoft Excel

ANALYSIS:

First, we load the Excel file and thoroughly go through the dataset and understand what information it is providing that is required for our project. What do you

see happening with the dataset? Once you have understood the problem, clean the data as necessary, and use your data analysis skills to explore the data set and derive insights to understand the important factors that make a movie more successful than others. So, we would like to analyse what kind of movies are more successful, famous directors and actors.

Our insights will help companies and investors make a successful movie with huge profits and with the best actor and director.

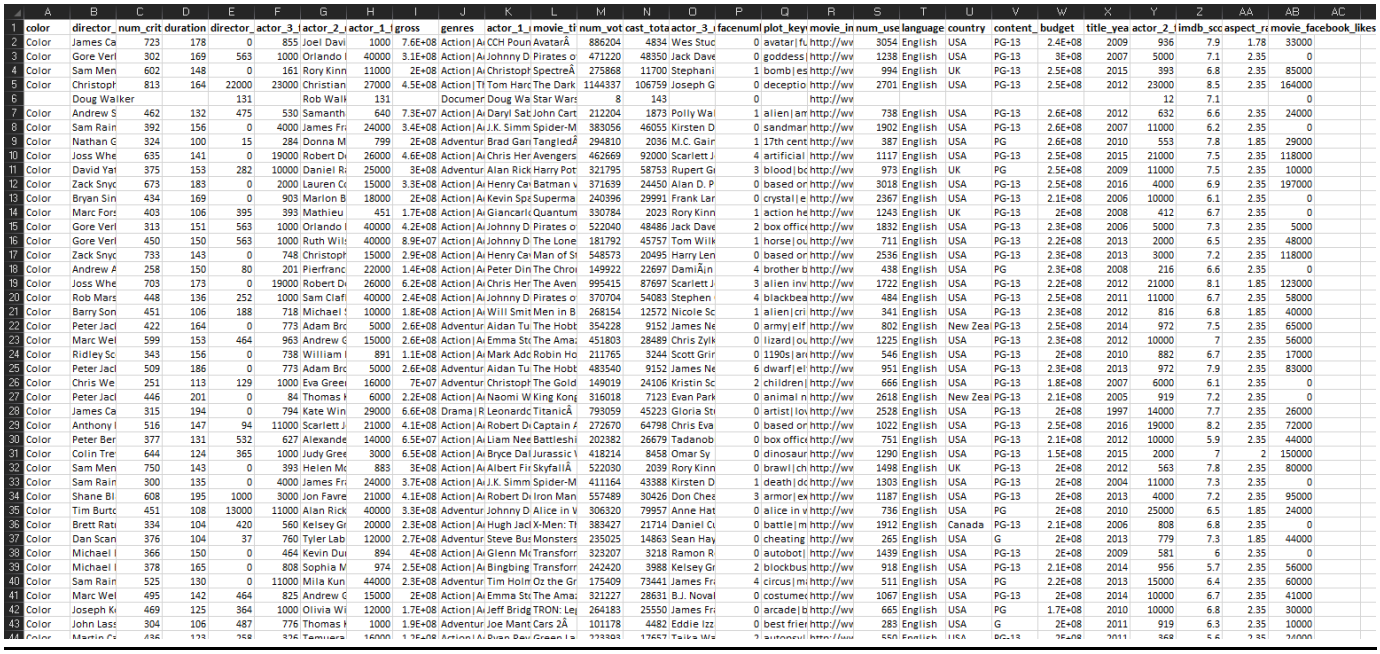
We can see that our dataset has the following variables:

movie_title	Title of the Movie
duration	Duration in minutes
director_name	Name of the Director of the Movie
director_facebook_likes	Number of likes of the Director on his Facebook Page
actor_1_name	Primary actor starring in the movie
actor_1_facebook_likes	Number of likes of the Actor_1 on his/her Facebook Page
actor_2_name	Another actor starring in the movie
actor_2_facebook_likes	Number of likes of the Actor_2 on his/her Facebook Page
actor_3_name	Another actor starring in the movie
actor_3_facebook_likes	Number of likes of the Actor_3 on his/her Facebook Page
num_user_for_reviews	Number of users who gave a review
num_critic_for_reviews	Number of critical reviews on imdb
num_voted_users	Number of people who voted for the movie
cast_total_facebook_likes	Total number of Facebook likes of the entire cast of the movie
movie_facebook_likes	Number of Facebook likes in the movie page
plot_keywords	Keywords describing the movie plot
facenumber_in_poster	Number of the actor who featured in the movie poster
color	Film colorization. 'Black and White' or 'Color'
genres	Film categorization like 'Animation', 'Comedy', 'Romance', 'Horror', 'Sci-Fi', 'Action', 'Family'
title_year	The year in which the movie is released (1916:2016)
language	English, Arabic, Chinese, French, German, Danish, Italian etc
country	Country where the movie is produced
content_rating	Content rating of the movie
aspect_ratio	Aspect ratio the movie was made in
movie_imdb_link	IMDB link of the movie
gross	Gross earnings of the movie in Dollars
budget	Budget of the movie in Dollars
imdb_score	IMDB Score of the movie on IMDB

INSIGHTS:

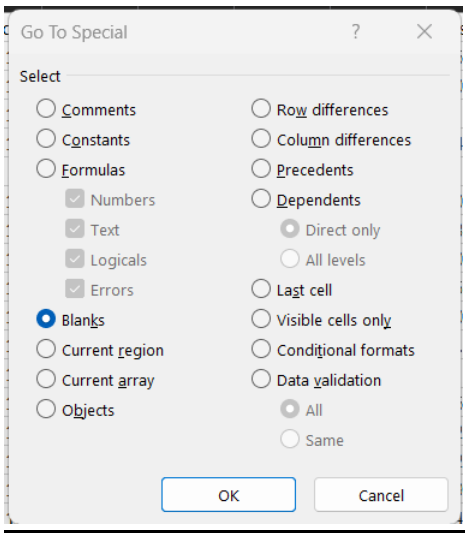
1.) CLEANING THE DATA: This is one of the most important steps to perform before moving forward with the analysis. This step involves preprocessing the data to make it suitable for analysis. It includes handling missing values, removing duplicates, converting data types if necessary, and possibly feature engineering.

RAW DATA

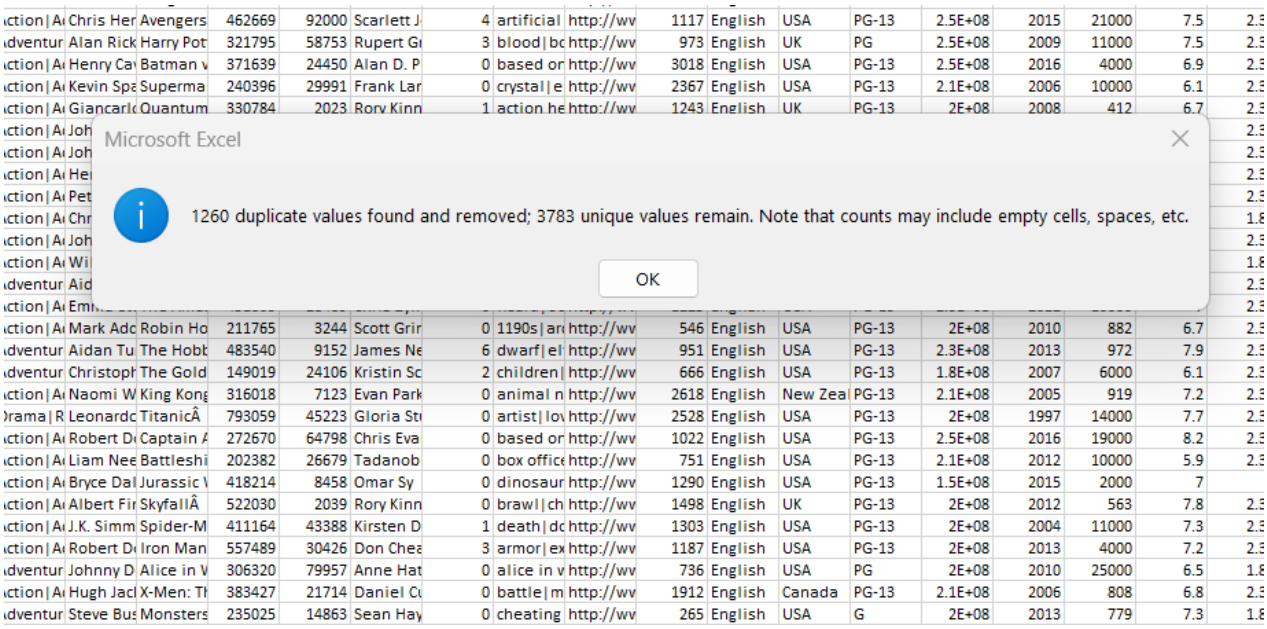


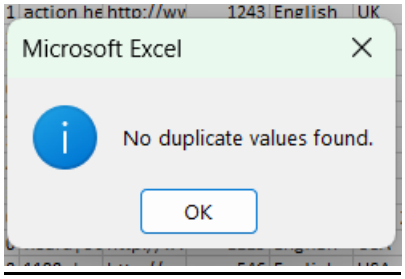
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	color	director_	num_crit	duration	director_	actor_3_f	actor_2_	actor_1_f	gross	genres	actor_1_	movie_ti	num_vot	cast_tota	actor_3_f	facenum	plot_key	movie_in	num_use	language	country	content_	budget	title_yea	actor_2_f	imdb_scc	aspect_ra	movie_facebook_likes	
2	Color	James Ca	723	178	0	855	Joel Davi	1000	7.6E+08	Action|Ai	CCH Poun	AvatarÂ	886204	4834	Wes Stuc	0	avatar|fu	http://ww	3054	English	USA	PG-13	2.4E+08	2009	936	7.9	1.78	33000	
3	Color	Gore Verl	302	169	563	1000	Orlando I	40000	3.1E+08	Action|Ai	Johnny D	Pirates o	471220	48350	Jack Dave	0	goddess|	http://ww	1238	English	USA	PG-13	3E+08	2007	5000	7.1	2.35	0	
4	Color	Sam Men	602	148	0	161	Rory Kinn	11000	2E+08	Action|Ai	Christoph	SpectreÂ	275868	11700	Stephani	1	bomb|es	http://ww	994	English	UK	PG-13	2.5E+08	2015	393	6.8	2.35	85000	
5	Color	Christoph	813	164	22000	23000	Christian	27000	4.5E+08	Action|Tl	Tom Hard	The Dark	1144337	106759	Joseph G	0	deceptio	http://ww	2701	English	USA	PG-13	2.5E+08	2012	23000	8.5	2.35	164000	
6		Doug Walker			131		Rob Walk	131		Documer	Doug Wa	Star Wars	8	143		0		http://ww							12	7.1		0	
7	Color	Andrew S	462	132	475	530	Samanth	640	7.3E+07	Action|Ai	Daryl Sat	John Cart	212204	1873	Polly Wal	1	alien|an	http://ww	738	English	USA	PG-13	2.6E+08	2012	632	6.6	2.35	24000	
8	Color	Sam Rain	392	156	0	4000	James Fri	24000	3.4E+08	Action|Ai	J.K. Simm	Spider-M	383056	46055	Kirsten D	0	sandmar	http://ww	1902	English	USA	PG-13	2.6E+08	2007	11000	6.2	2.35	0	
9	Color	Nathan G	324	100	15	284	Donna M	799	2E+08	Adventur	Brad Gari	TangledÂ	294810	2036	M.C. Gain	1	17th cent	http://ww	387	English	USA	PG	2.6E+08	2010	553	7.8	1.85	29000	
10	Color	Joss Whe	635	141	0	19000	Robert D	26000	4.6E+08	Action|Ai	Chris Her	Avengers	462669	92000	Scarlett J	4	artificial	http://ww	1117	English	USA	PG-13	2.5E+08	2015	21000	7.5	2.35	118000	
11	Color	David Yat	375	153	282	10000	Daniel Ra	25000	3E+08	Adventur	Alan Rick	Harry Pot	321795	58753	Rupert Gi	3	blood|bc	http://ww	973	English	UK	PG	2.5E+08	2009	11000	7.5	2.35	10000	
12	Color	Zack Snyc	673	183	0	2000	Lauren Cc	15000	3.3E+08	Action|Ai	Henry Cav	Batman v	371639	24450	Alan D. P	0	based or	http://ww	3018	English	USA	PG-13	2.5E+08	2016	4000	6.9	2.35	197000	
13	Color	Bryan Sin	434	169	0	903	Marlon B	18000	2E+08	Action|Ai	Kevin Spa	Superma	240396	29991	Frank Lar	0	crystal|e	http://ww	2367	English	USA	PG-13	2.1E+08	2006	10000	6.1	2.35	0	
14	Color	Marc Fors	403	106	395	393	Mathieu	451	1.7E+08	Action|Ai	Giancarlo	Quantum	330784	2023	Rory Kinn	1	action he	http://ww	1243	English	UK	PG-13	2E+08	2008	412	6.7	2.35	0	
15	Color	Gore Verl	313	151	563	1000	Orlando I	40000	4.2E+08	Action|Ai	Johnny D	Pirates o	522040	48486	Jack Dave	2	box offic	http://ww	1832	English	USA	PG-13	2.3E+08	2006	5000	7.3	2.35	5000	
16	Color	Gore Verl	450	150	563	1000	Ruth Wils	40000	8.9E+07	Action|Ai	Johnny D	The Lone	181792	45757	Tom Wilk	1	horse|ou	http://ww	711	English	USA	PG-13	2.2E+08	2013	2000	6.5	2.35	48000	
17	Color	Zack Snyc	733	143	0	748	Christoph	15000	2.9E+08	Action|Ai	Henry Cav	Man of St	548573	20495	Harry Len	0	based or	http://ww	2536	English	USA	PG-13	2.3E+08	2013	3000	7.2	2.35	118000	
18	Color	Andrew A	258	150	80	201	Pierfranc	22000	1.4E+08	Action|Ai	Peter Din	The Chro	149922	22697	DamiÃ¡n	4	brother b	http://ww	438	English	USA	PG	2.3E+08	2008	216	6.6	2.35	0	
19	Color	Joss Whe	703	173	0	19000	Robert D	26000	6.2E+08	Action|Ai	Chris Her	The Aven	995415	87697	Scarlett J	3	alien inv	http://ww	1722	English	USA	PG-13	2.2E+08	2012	21000	8.1	1.85	123000	
20	Color	Rob Mars	448	136	252	1000	Sam Clafl	40000	2.4E+08	Action|Ai	Johnny D	Pirates o	370704	54083	Stephen	4	blackbea	http://ww	484	English	USA	PG-13	2.5E+08	2011	11000	6.7	2.35	58000	
21	Color	Barry Son	451	106	188	718	Michael S	10000	1.8E+08	Action|Ai	Will Smit	Men in B	268154	12572	Nicole Sc	1	alien|cri	http://ww	341	English	USA	PG-13	2.3E+08	2012	816	6.8	1.85	40000	
22	Color	Peter Jacl	422	164	0	773	Adam Brc	5000	2.6E+08	Adventur	Aidan Tu	The Hobb	354228	9152	James Ne	0	army|elf	http://ww	802	English	New Zeal	PG-13	2.5E+08	2014	972	7.5	2.35	65000	
23	Color	Marc Wel	599	153	464	963	Andrew G	15000	2.6E+08	Action|Ai	Emma Sto	The Amaz	451803	28489	Chris Zylk	0	lizard|ou	http://ww	1225	English	USA	PG-13	2.3E+08	2012	10000	7	2.35	56000	
24	Color	Ridley Sc	343	156	0	738	William I	891	1.1E+08	Action|Ai	Mark Adc	Robin Ho	211765	3244	Scott Grir	0	1190s|ar	http://ww	546	English	USA	PG-13	2E+08	2010	882	6.7	2.35	17000	
25	Color	Peter Jacl	509	186	0	773	Adam Brc	5000	2.6E+08	Adventur	Aidan Tu	The Hobb	483540	9152	James Ne	6	dwarf|el	http://ww	951	English	USA	PG-13	2.3E+08	2013	972	7.9	2.35	83000	
26	Color	Chris We	251	113	129	1000	Eva Gree	16000	7E+07	Adventur	Christoph	The Gold	149019	24106	Kristin Sc	2	children|	http://ww	666	English	USA	PG-13	1.8E+08	2007	6000	6.1	2.35	0	
27	Color	Peter Jacl	446	201	0	84	Thomas H	6000	2.2E+08	Action|Ai	Naomi W	King Kong	316018	7123	Evan Park	0	animal n	http://ww	2618	English	New Zeal	PG-13	2.1E+08	2005	919	7.2	2.35	0	
28	Color	James Ca	315	194	0	794	Kate Win	29000	6.6E+08	Drama|R	Leonardc	TitanicÂ	793059	45223	Gloria St	0	artist|lo	http://ww	2528	English	USA	PG-13	2E+08	1997	14000	7.7	2.35	26000	
29	Color	Anthony I	516	147	94	11000	Scarlett J	21000	4.1E+08	Action|Ai	Robert Dc	Captain A	272670	64798	Chris Eva	0	based or	http://ww	1022	English	USA	PG-13	2.5E+08	2016	19000	8.2	2.35	72000	
30	Color	Peter Ber	377	131	532	627	Alexande	14000	6.5E+07	Action|Ai	Liam Nee	Battleshi	202382	26679	Tadanob	0	box offic	http://ww	751	English	USA	PG-13	2.1E+08	2012	10000	5.9	2.35	44000	
31	Color	Colin Tre	644	124	365	1000	Judy Gree	3000	6.5E+08	Action|Ai	Bryce Dal	Jurassic V	418214	8458	Omar Sy	0	dinosaur	http://ww	1290	English	USA	PG-13	1.5E+08	2015	2000	7	2	150000	
32	Color	Sam Men	750	143	0	393	Helen Mc	883	3E+08	Action|Ai	Albert Fir	SkyfallÂ	522030	2039	Rory Kinn	0	brawl|ch	http://ww	1498	English	UK	PG-13	2E+08	2012	563	7.8	2.35	80000	
33	Color	Sam Rain	300	135	0	4000	James Fri	24000	3.7E+08	Action|Ai	J.K. Simm	Spider-M	411164	43388	Kirsten D	1	death|dc	http://ww	1303	English	USA	PG-13	2E+08	2004	11000	7.3	2.35	0	
34	Color	Shane Bl	608	195	1000	3000	Jon Favre	21000	4.1E+08	Action|Ai	Robert Dc	Iron Man	557489	30426	Don Chee	3	armor|ex	http://ww	1187	English	USA	PG-13	2E+08	2013	4000	7.2	2.35	95000	
35	Color	Tim Burtc	451	108	13000	11000	Alan Rick	40000	3.3E+08	Adventur	Johnny D	Alice in W	306320	79957	Anne Hat	0	alice in v	http://ww	736	English	USA	PG	2E+08	2010	25000	6.5	1.85	24000	
36	Color	Brett Rat	334	104	420	560	Kelsey Gr	20000	2.3E+08	Action|Ai	Hugh Jacl	X-Men: Th	383427	21714	Daniel Cu	0	battle|m	http://ww	1912	English	Canada	PG-13	2.1E+08	2006	808	6.8	2.35	0	
37	Color	Dan Scan	376	104	37	760	Tyler Lab	12000	2.7E+08	Adventur	Steve Bus	Monsters	235025	14863	Sean Hay	0	cheating	http://ww	265	English	USA	G	2E+08	2013	779	7.3	1.85	44000	
38	Color	Michael I	366	150	0	464	Kevin Dur	894	4E+08	Action|Ai	Glenn Mc	Transforr	323207	3218	Ramon R	0	autobot|	http://ww	1439	English	USA	PG-13	2E+08	2009	581	6	2.35	0	
39	Color	Michael I	378	165	0	808	Sophia M	974	2.5E+08	Action|Ai	Bingbing	Transforr	242420	3988	Kelsey Gr	2	blockbus	http://ww	918	English	USA	PG-13	2.1E+08	2014	956	5.7	2.35	56000	
40	Color	Sam Rain	525	130	0	11000	Mila Kun	44000	2.3E+08	Adventur	Tim Holm	Oz the Gr	175409	73441	James Fri	4	circus|m	http://ww	511	English	USA	PG	2.2E+08	2013	15000	6.4	2.35	60000	
41	Color	Marc Wel	495	142	464	825	Andrew G	15000	2E+08	Action|Ai	Emma Sto	The Amaz	321227	28631	B.J. Noval	0	costume	http://ww	1067	English	USA	PG-13	2E+08	2014	10000	6.7	2.35	41000	
42	Color	Joseph K	469	125	364	1000	Olivia Wi	12000	1.7E+08	Action|Ai	Jeff Bridg	TRON: Le	264183	25550	James Fri	0	arcade|b	http://ww	665	English	USA	PG	1.7E+08	2010	10000	6.8	2.35	30000	
43	Color	John Lass	304	106	487	776	Thomas H	1000	1.9E+08	Adventur	Joe Mant	Cars 2Â	101178	4482	Eddie Izz	0	best frier	http://ww	283	English	USA	G	2E+08	2011	919	6.3	2.35	10000	
44	Color	Martin Cs	436	123	258	326	Temuera	16000	1.2E+08	Action|Ai	Ryan Rey	Green La	223393	17657	Talka Wa	2	autopsy|	http://ww	550	English	USA	PG-13	2E+08	2011	368	5.6	2.35	24000	

- Removing blank values:



- Removing duplicates:
  Deleted all the 1260 duplicate records from the table, rows were reduced from 5043 to 3783.





TASKS:

A.) 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 the IMDB scores' descriptive statistics (mean, median, mode, range, variance, standard deviation).

Total Genres	Genres_Name	Mean	Mode	median	variance	stdev	Max	Min
951	Action	6.29	6.1	6.3	1.064	1.031626	9	2.1
366	Adventure	6.56	7.3	6.7	1.259	1.122021	8.6	2.3
45	Animation	6.74	7.1	7	0.922	0.959954	8	4.5
204	Biography	7.16	7	7.2	0.483	0.695060	8.9	4.5
984	Comedy	6.17	6.4	6.3	1.054	1.026719	8.8	1.9
253	Crime	6.94	7.4	7	0.755	0.868749	9.3	3.3
26	Documentary	6.80	7.5	7.45	2.833	1.683285	8.5	1.6
659	Drama	6.83	6.7	6.9	0.823	0.907071	8.8	2.1
3	Family	6.50	0	5.9	0.987	0.993311	7.9	5.7
37	Fantasy	6.28	6.8	6.5	0.778	0.881901	7.9	4.3
159	Horror	5.84	5.9	5.9	1.053	1.026221	8.5	2.3
2	Musical	6.75	0	6.75	0.203	0.450000	7.2	6.3
23	Mystery	6.65	7.1	6.7	1.142	1.068469	8.5	3.3
1	Romance	7.10	0	7.1	0.000	0.000000	7.1	7.1
7	Sci-Fi	6.63	0	6.4	1.051	1.024994	8.2	5
1	Thriller	4.80	0	4.8	0.000	0.000000	4.8	4.8
2	Western	8.10	0	8.1	0.640	0.800000	8.9	7.3

INSIGHTS:

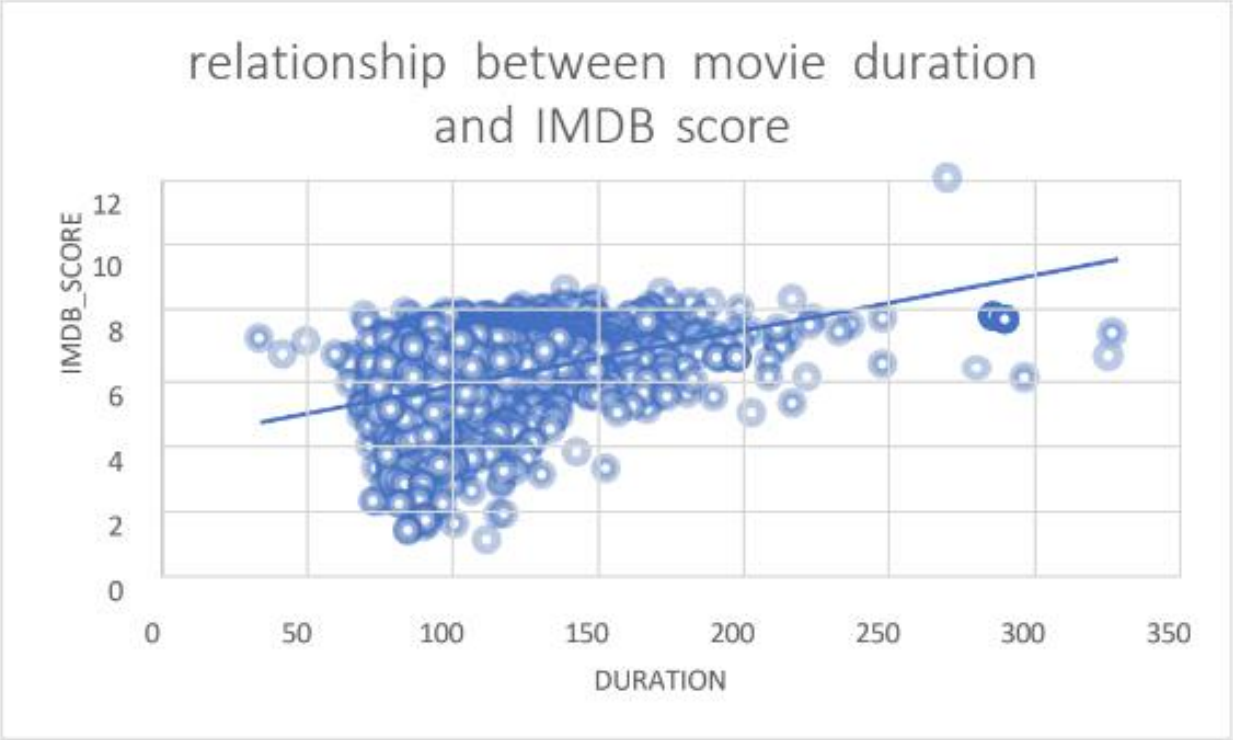
- **Most Common Genres**: The analysis shows the frequency of each genre in the dataset. For For example, there are 951 action movies, making it the most common genre in the dataset.
- **Average Ratings**: The mean (average) IMDB score for each genre indicates the overall rating for movies in that genre. For example, Drama movies have an average rating of 6.83, while Horror movies have an average rating of 5.84.

- **Central Tendency**: The mode represents the most frequently occurring IMDB score in each genre, while the median represents the middle value. These measures help understand the typical rating for movies in each genre. For example, the mode for Action movies is 6.1, while the median is 6.3.
- **Variability:** The range (Max-Min) shows the difference between the highest and lowest IMDB scores in each genre, indicate the variability of ratings. For example, Adventure movies have a range of 8.6 - 2.3 = 6.3, suggesting a wide range of ratings.
- **Overall Rating Distribution**: By comparing the mean, median, and mode, you can understand the distribution of ratings within each genre. For example, Comedy movies have a mean of 6.17, close to the median of 6.3, suggesting a relatively symmetric distribution of ratings.

These insights help understand the distribution and variability of IMDB scores for different movie genres, providing valuable information for further analysis and decision-making.

B.) MOVIE DURATION ANALYSIS: Analyse the distribution of movie durations and its impact on the IMDB score.

#Task: Analyse the distribution of movie durations and identify the relationship between movie duration and IMDB score.



Mean	Median	Stdev
6.465673	6.6	1.053644

INSIGHTS:

- Mean IMDB Score: The average IMDB score across the movies in your dataset is approximately 6.47.

- Median IMDB Score: The median IMDB score is 6.6, which represents the middle value when all scores are sorted in ascending order.
- Standard Deviation (Stdev): The standard deviation of IMDB scores is approximately 1.05. This indicates the variability or spread of scores around the mean.

Now, let's interpret the plot:

The scatter plot shows a positive correlation between movie duration (x-axis) and IMDB score

(y-axis).

As movie duration increases, there seems to be a tendency for IMDB scores to also rise.

Longer movies tend to receive higher IMDB ratings, suggesting audiences appreciate well-crafted, immersive storytelling in lengthier films.

C.) LANGUAGE ANALYSIS: examine the distribution of movies based on their language.

#Task: Determine the most common languages used in movies and analyse their impact on the IMDB score using descriptive statistics.

Unique_Language	Total_Count_languages	Mean	Stdev	Median
Aboriginal	2	9.25	0.070710678	8
Arabic	1	9	0.2	8
Aramaic	1	9	0	8.4
Bosnian	1	8.9	0.2	7.3
Cantonese	7	8.842857143	0.053452248	7.2
Czech	1	8.8	0	7.1
Danish	3	8.7	0	7.6
Dari	2	8.7	0	7.5
Dutch	3	8.666666667	0.057735027	8.1
English	3566	6.487464947	0.989916863	6.5
Filipino	1	4.3	0.3	7.9
French	34	5.588235294	1.436275635	7.2
German	10	6.39	1.110005005	7.1
Hebrew	1	3.9	1.192895637	7.7
Hindi	5	5.14	1.192895637	7.6
Hungarian	1	3.8	1.839232932	4.3
Indonesian	2	4.2	0.565685425	7.3
Italian	7	6.557142857	1.411685922	6.5
Japanese	10	5.35	1.839232932	7.6
Kazakh	1	4.7	0	7.8
Korean	5	6.16	1.499333185	7.2
Mandarin	14	5.9	1.48531271	7
Maya	1	5.5	0.7	7.2
Mongolian	1	3.3	0	7.1
None	1	3.9	0	2.5
Norwegian	4	6.125	1.912023361	7.5
Persian	3	4.466666667	2.800595175	7.6
Portuguese	5	5.02	2.256546033	0.8
Romanian	1	6.3	0.2	0.2
Russian	1	2.7	0.4	6.6
Spanish	23	5.439130435	2.098427795	7.2
Thai	3	5.433333333	3.074627349	7.4
Vietnamese	1	1.6	0	7.1
Zulu	1	6.6	0	6.9

INSIGHTS:

- **Number of Movies:** The analysis shows the total count of movies for each language. For example, English is the most common language with 3566 movies, followed by French with 34 movies.
- **Average IMDB Score:** The mean IMDB score for movies in each language indicates the overall rating for movies in that language. For example, movies in Danish have an average rating of 8.7, while movies in Russian have an average rating of 2.7.
- **Standard Deviation:** The standard deviation of IMDB scores for movies in each language indicates the variability of ratings. A higher standard deviation suggests more variability in ratings. For example, movies in Hungarian have a high standard deviation of 1.84, indicating a wide range of ratings.
- **Impact of Language on IMDB Score:** By comparing the average IMDB scores and standard deviations across languages, you can analyse the impact of language on the IMDB score. For example, movies in English have a mean score of 6.49 with a standard deviation of 0.99, while movies in Spanish have a mean score of 5.44 with a standard deviation of 2.10.

- **Distribution of Ratings:** The median IMDB score for each language indicates the central tendency of ratings. For example, the median IMDB score for movies in Mandarin is 7, suggesting that the ratings are evenly distributed around this value.

#Overall, the analysis provides insights into the distribution of movies across different languages and their impact on IMDB scores, helping to understand the preferences and ratings of movies in different languages.

D.) **DIRECTOR ANALYSIS:** Influence of directors on movie ratings.

#Task: Identify the top directors based on their average IMDB score and analyse their contribution to the success of movies using percentile calculations.

Top_director_Name	Average of imdb_score	Percentile
Akira Kurosawa	8.7	8.1
Alfred Hitchcock	8.5	8.1
Asghar Farhadi	8.4	8.106666667
Billy Wilder	8.3	8.137222222
Charles Chaplin	8.6	8.163333333
Christopher Nolan	8.425	8.2
Damien Chazelle	8.5	8.2
David Sington	8.1	8.2
Elia Kazan	8.2	8.2
Fritz Lang	8.3	8.210833333
George Roy Hill	8.2	8.228888889
Hayao Miyazaki	8.225	8.266666667
Je-kyu Kang	8.1	8.3
Joshua Oppenheimer	8.2	8.3
Juan JosÃ© Campanella	8.2	8.3
Lee Unkrich	8.3	8.363333333
Lenny Abrahamson	8.3	8.4
Majid Majidi	8.5	8.4175
Michael Wadleigh	8.1	8.431111111
Milos Forman	8.133333333	8.484444444
Pete Docter	8.233333333	8.5
Quentin Tarantino	8.2	8.5
Richard Marquand	8.4	8.5
Ron Fricke	8.5	8.59
Sergio Leone	8.433333333	8.6
Terry George	8.1	8.696666667
Tim Miller	8.1	8.1
Tony Kaye	8.6	8.1
Victor Fleming	8.15	8.2
William Wyler	8.1	8.6
Grand Total	**8.292727273**	

INSIGHTS:

- Akira Kurosawa (Average IMDb Score: 8.7):

  Kurosawa is renowned for his masterful storytelling and influential films. His high average IMDb score reflects the impact of classics like “Seven Samurai” and “Rashomon.”

- Alfred Hitchcock (Average IMDb Score: 8.5):

  Hitchcock’s suspenseful thrillers, such as “Psycho” and “Rear Window,” have left an indelible mark on cinema. His consistently high ratings attest to his enduring legacy.

- Asghar Farhadi (Average IMDb Score: 8.4):

  Farhadi, an Iranian director, gained a claim for his emotionally charged dramas like “A Separation.” His IMDb score places him among the elite.

- Billy Wilder (Average IMDb Score: 8.3):

  Wilder’s wit and versatility shine in films like “Sunset Boulevard” and “Some Like It Hot.” His percentile suggests sustained excellence.

- Charles Chaplin (Average IMDb Score: 8.6):

  Chaplin’s silent comedies, including “City Lights” and “Modern Times,” continue to resonate. His high percentile reflects their timeless appeal.

- Christopher Nolan (Average IMDb Score: 8.425):
  Nolan's mind-bending narratives, like "Inception" and "The Dark Knight," captivate audiences. His percentile indicates consistent quality.
- Hayao Miyazaki (Average IMDb Score: 8.225):

  Miyazaki's animated works, such as "Spirited Away" and "My Neighbor Totoro," enchant viewers. His percentile underscores their global impact.

- Quentin Tarantino (Average IMDb Score: 8.2):

  Tarantino's unique style, seen in films like "Pulp Fiction" and "Kill Bill," resonates with cinephiles. His percentile reflects his cult following.

- Sergio Leone (Average IMDb Score: 8.433):

  Leone's epic spaghetti westerns, including "The Good, the Bad and the Ugly," remain iconic. His percentile highlights their enduring popularity.

- Milos Forman (Average IMDb Score: 8.133):

  Forman's diverse filmography, spanning "One Flew Over the Cuckoo's Nest" to "Amadeus," contributes to his respectable percentile. Keep in mind that this analysis is based on the provided data, and other factors not considered here may also influence IMDB scores. However, the positive trend observed in the plot aligns with the intuition that longer

movies often offer more depth and engagement for viewers.

E.) <u>BUDGET ANALYSIS:</u> Explore the relationship between movie budgets and their financial success.

#Task: Analyse the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Total_Profit	Corelation_Cofficient	Highest Profit Margin
3341469	0.098318102	Avatar - Profit: 523505847
128821952		
348316061		
44300000		
283019252		
74067179		
99930000		
4900000		
132568851		
220837577		
-25976605		
274691196		
272158751		
246478898		
108383253		
21836394		
449935665		
107600000		
4263397		
-1730939		
22991439		
146119491		
67125340		

INSIGHTS:

Correlation Coefficient (0.098318102):

The correlation coefficient measures the strength and direction of the linear relationship between movie budgets and gross earnings.

In this case, the positive correlation coefficient suggests a weak positive relationship. As budgets increase, gross earnings tend to increase slightly, but not significantly.

Highest Profit Margin Movie: “Avatar”:

- The movie “Avatar” achieved the highest profit margin. Its profit is 523,505,847.
- The profit margin is calculated as follows:

Profit Margin = (Gross Earnings - Budget)

The high-profit margin indicates that “Avatar” was highly successful financially.

- Other Movies:

The remaining movies have varying profit margins, some positive and some negative. Positive profit margins indicate profitability, while negative margins imply losses.

Further analysis would require additional context, such as each movie's specific budget and gross earnings.

while the overall correlation between budgets and gross earnings is weak, “Avatar” stands out as a blockbuster with an impressive profit margin.

CONCLUSION:

- English Dominance: English is the most common language used in movies, with a significantly higher number of movies than other languages. This indicates the dominance of English-language cinema in the dataset.
- IMDB Scores: The average IMDB scores vary across different languages, with some languages like Danish and Norwegian having relatively high average scores, while others like Russian and Hungarian have lower average scores.
- Variability in Ratings: There is variability in IMDB scores within each language, as indicated by the standard deviations. Languages with higher standard deviations, such as Portuguese and Thai, have more variability in ratings compared to languages with lower standard deviations.
- Impact of Language: The language of a movie appears to have an impact on its IMDB score, as evidenced by the differences in average scores across languages. However, other factors such as the quality of the

movie, its storyline, and the actors' performance may also play a significant role in determining the IMDB score.

#Further Analysis:

Further analysis could involve exploring the relationship between language and other factors such as genre, duration, and production budget to gain a deeper understanding of the factors that contribute to the success of a movie.

Overall, the analysis provides valuable insights into the distribution of movies across different languages and their impact on IMDB scores, highlighting the diversity and complexity of the movie industry.

-end-

HYPERLINK TO EXCEL FILE:

https://docs.google.com/spreadsheets/d/1w0bEWI0g-FiGfRob9d9FG325D2amUdRU/edit?usp=sharing&ouid=102683227032029211056&rtpof=true&sd=true

HYPERLINK TO PPT:

https://drive.google.com/file/d/11KVPEFZmB7c9tDH82WM6fyQ3kBdiIV5l/view?usp=sharing

HYPERLINK TO VIDEO SUBMISSION:

https://drive.google.com/file/d/1NJGVbqEDrNl2CHO3_J4kyfrxuzJtAhy_/view?usp=sharing