

IMDB Movie Analysis

- Project Description: We have been given data about movies with imdb ratings and all other information about movies , our role as a data analyst is to figure out “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.
- IMDB:
IMDb stands for the Internet Movie Database. It is an online database that provides information about films, television programs, video games, and streaming content

DATA CLEANING & DATA ANALYSIS

- ❖ Data cleaning, also known as data cleansing or data scrubbing, is the process of identifying and correcting errors or inconsistencies in a dataset. The goal is to improve the quality of the data, making it accurate, reliable, and ready for analysis.
- ❖ Data analysis is the process of inspecting, cleaning, transforming, and modeling data to discover useful information, draw conclusions, and support decision-making. It involves examining raw data with the goal of extracting meaningful insights that can inform business decisions, research findings, or other applications

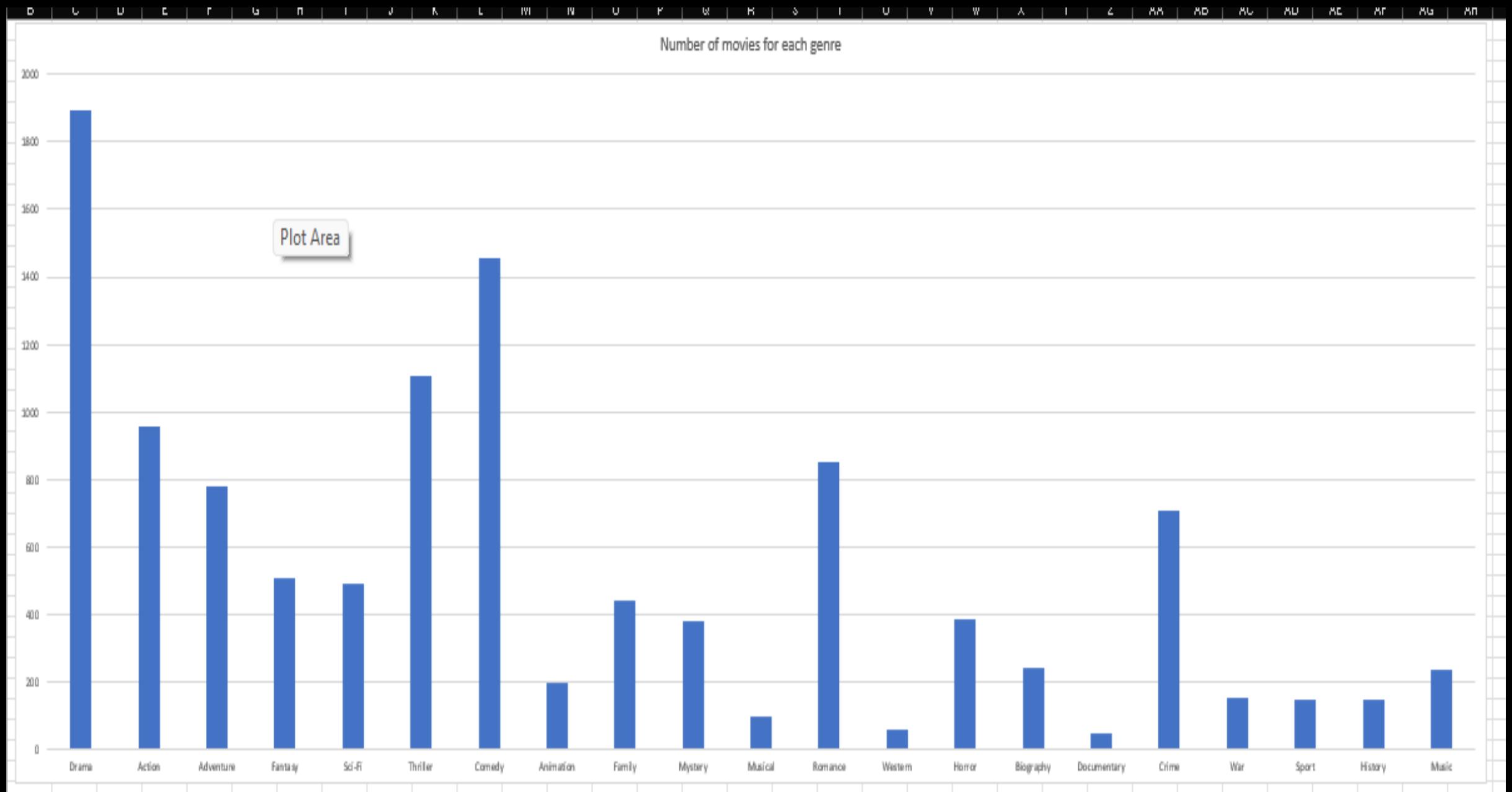
TASK A:Movie Genre Analysis

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

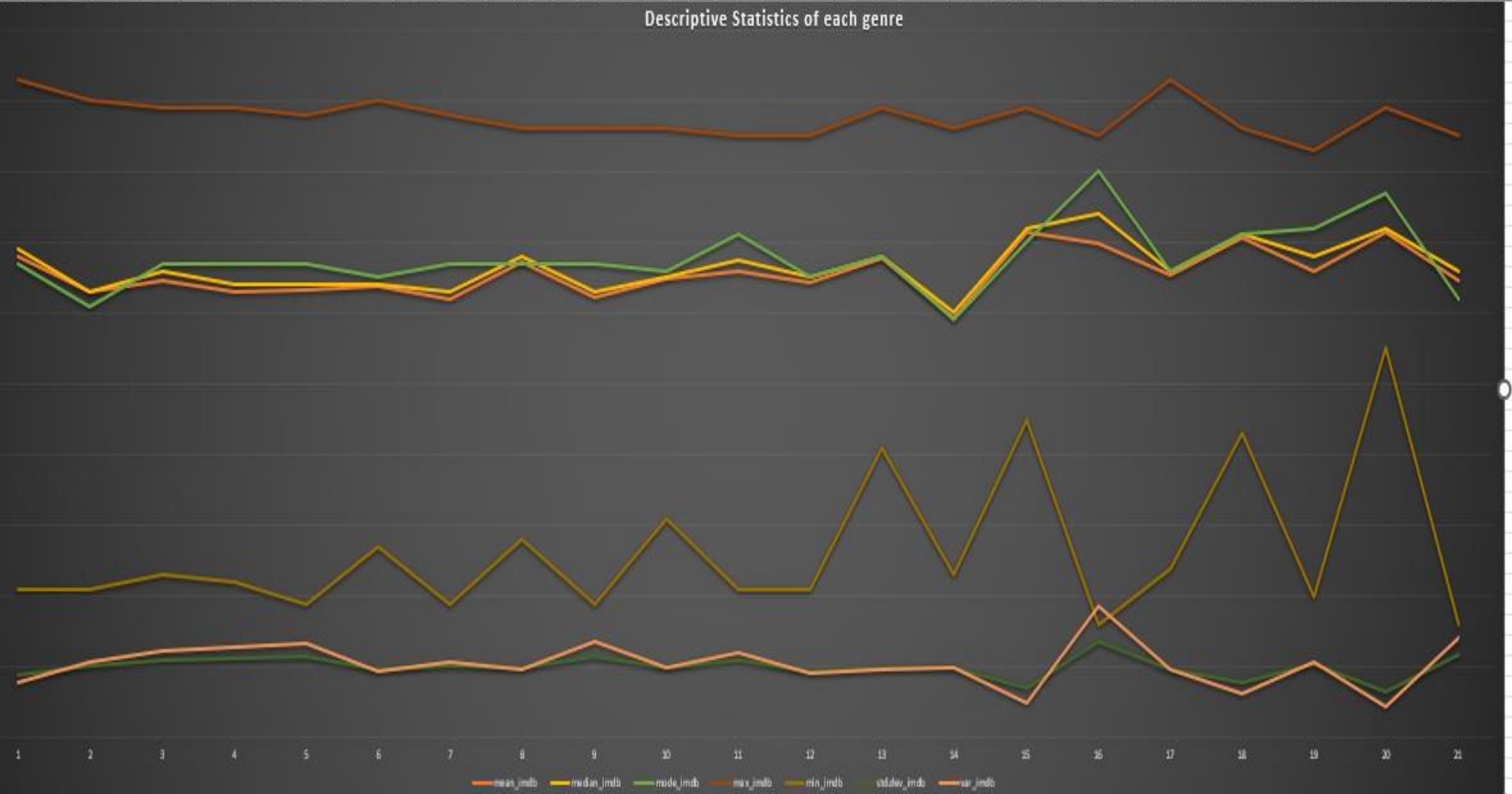
- ❖ Task: Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

Most Common genres of movies and its descriptive statistics

		C	D	E	F	G	H	I	J	K
	genre(seperated)	count of movies	mean_imdb	median_imdb	mode_imdb	max_imdb	min_imdb	std.dev_imdb	var_imdb	
	Drama	1893	6.787955626	6.9	6.7	9.3	2.1	0.892738029	0.796981188	
	Action	954	6.288888889	6.3	6.1	9	2.1	1.031476432	1.063943629	
	Adventure	777	6.454311454	6.6	6.7	8.9	2.3	1.110525318	1.233266482	
	Fantasy	508	6.29015748	6.4	6.7	8.9	2.2	1.128715027	1.273997613	
	Sci-Fi	493	6.325760649	6.4	6.7	8.8	1.9	1.155744836	1.335746125	
	Thriller	1108	6.376263538	6.4	6.5	9	2.7	0.966695276	0.934499757	
Romance	Comedy	1458	6.186351166	6.3	6.7	8.8	1.9	1.032423065	1.065897386	
	Animation	197	6.708121827	6.8	6.7	8.6	2.8	0.987554548	0.975263985	
	Family	442	6.216515837	6.3	6.7	8.6	1.9	1.164471638	1.355994195	
	Mystery	380	6.476578947	6.5	6.6	8.6	3.1	1.001316965	1.002635665	
	Musical	96	6.596875	6.75	7.1	8.5	2.1	1.096153609	1.201552734	
	Romance	853	6.434232122	6.5	6.5	8.5	2.1	0.955627757	0.91322441	
	Western	58	6.765517241	6.8	6.8	8.9	4.1	0.989871417	0.979845422	
	Horror	387	5.923255814	6	5.9	8.6	2.3	0.994624697	0.989278289	
	Biography	241	7.142323651	7.2	7	8.9	4.5	0.71005902	0.504183812	
	Documentary	48	6.975	7.4	8	8.5	1.6	1.361601386	1.853958333	
	Crime	707	6.541301273	6.6	6.6	9.3	2.4	0.982285822	0.964885435	
	War	152	7.059868421	7.1	7.1	8.6	4.3	0.79722049	0.635560509	
	Sport	147	6.589115646	6.8	7.2	8.3	2	1.040481917	1.082602619	
	History	148	7.150675676	7.2	7.7	8.9	5.5	0.668246659	0.446553598	
	Music	233	6.457081545	6.6	6.2	8.5	1.6	1.18483047	1.403823242	



Descriptive Statistics of each genre



Report and Data Story

- ❖ The drama genre has the highest number of movies among all genres , even though its ratings average is not very high.
- ❖ Drama genre has hit one of the highest rating among all genres that's why everyone is preferring drama genre.
- ❖ War, History, Biography, Documentary are the genres which has the best average among all genres it is very safe to make war, history, biography genre movies because its minimum average is also more than 4 which is not bad compare to others.
- ❖ There are many movies with action,comedy,horror,music and thriller genre but according to data the average rating is not very high which tells the story that the movies didn't attracted the audience so there is also scope in these genres, we must attract the audience by finding the loopholes of others and trying to give the best quality movie.

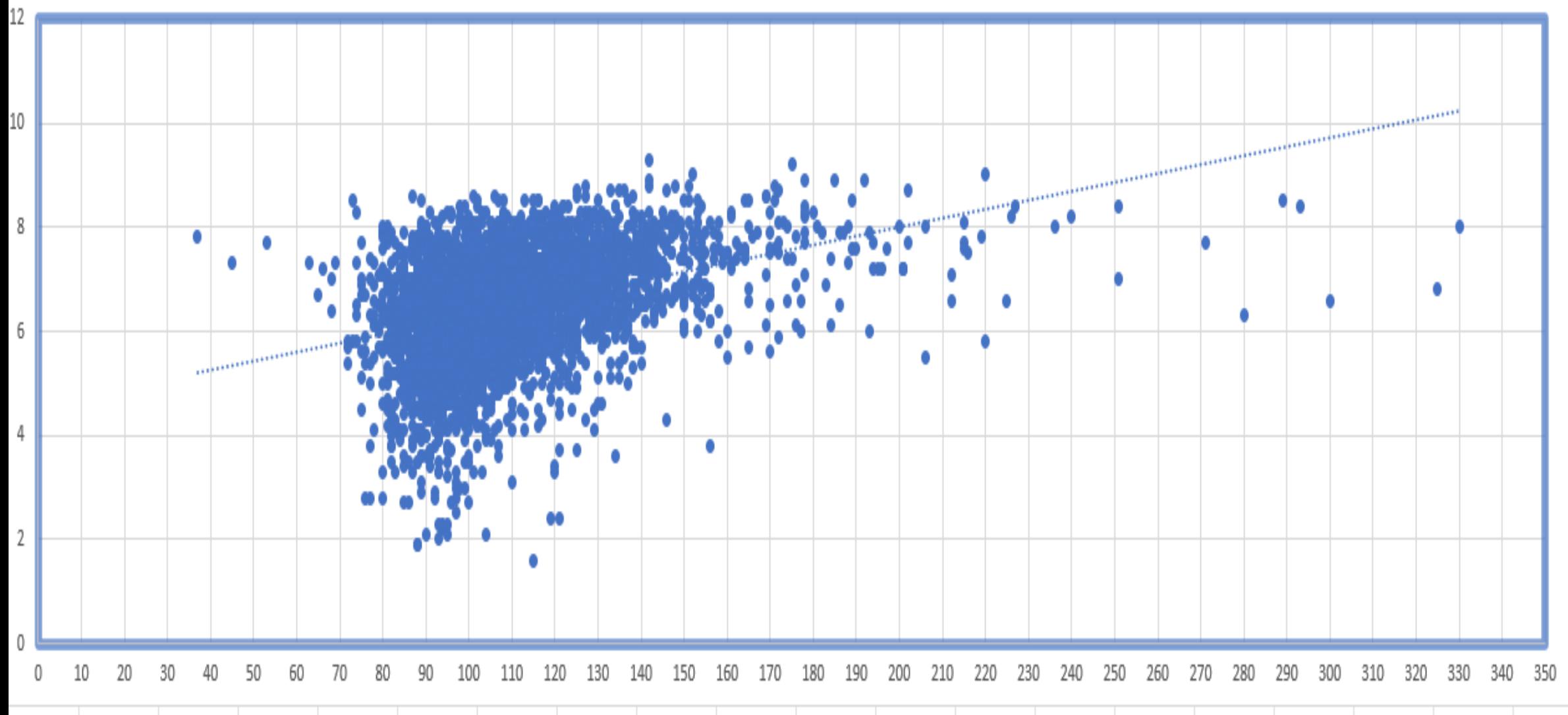
Task B: Movie Duration Analysis

- ❖ 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.

Movie duration based on Imdb score

Movie Duration	
Mean	110.2131
Median	106
Mode	101
Standard Deviation	22.67561
Variance	514.1832

Relationship between imdb_score & movie duration



Report & Data Story

- ❖ The average duration of movies is 110 mins
- ❖ There is lot of movies which vary from average movie duration as a result variance is very high
- ❖ Most of the movies are in the range of 80 –140 , which is nearer to average duration of the movie.
- ❖ There are movies which also has very high duration which is above 150-400 which is very abnormal and also not good

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.

Languges	count of movies	mean_imdb	median_imdb	Std.dev_imdb	var_imdb	
English	3591	6.4265107	6.5	1.048170912	1.09866	
Aboriginal	2	6.95	6.95	0.55	0.3025	
Arabic	1	7.2	7.2	0	0	
Aramaic	1	7.1	7.1	0	0	
Bosnian	1	4.3	4.3	0	0	
Cantonese	7	7.3428571	7.3	0.324509048	0.10531	
Czech	1	7.4	7.4	0	0	
Danish	3	7.9	8.1	0.43204938	0.18667	
Dari	2	7.08125	7.4	0.709065186	0.50277	
Dutch	3	7.5666667	7.8	0.329983165	0.10889	
Filipino	1	6.7	6.7	0	0	
French	34	7.3558824	7.3	0.51173935	0.26188	
German	11	7.7636364	7.8	0.644237008	0.41504	
Hebrew	1	8	8	0	0	
Hindi	5	7.22	7.4	0.716658915	0.5136	
Hungarian	1	7.1	7.1	0	0	

Indonesian	2	7.9	7.9	0.3	0.09
Italian	7	7.1857143	7	1.069617517	1.14408
Japanese	10	7.66	8	0.939361485	0.8824
Kazakh	1	6	6	0	0
Korean	5	7.7	7.7	0.509901951	0.26
Mandarin	14	7.0214286	7.25	0.737930089	0.54454
Maya	1	7.8	7.8	0	0
Mongolian	1	7.3	7.3	0	0
None	1	8.5	8.5	0	0
Norwegian	4	7.15	7.3	0.497493719	0.2475
Persian	3	8.1333333	8.4	0.449691252	0.20222
Portuguese	5	7.76	8	0.875442745	0.7664
Romanian	1	7.9	7.9	0	0
Russian	1	6.5	6.5	0	0
Spanish	23	7.0826087	7.2	0.841660974	0.70839
Thai	3	6.6333333	6.6	0.368178701	0.13556
Vietnamese	1	7.4	7.4	0	0
Zulu	1	7.3	7.3	0	0

Report & Data Story

- ❖ Most common language used in movies is English
- ❖ But Imdb ratings average of English movies is not high as compared to other languages
- ❖ Imdb ratings of movies in persian & Hebrew language is high when compared to others but the number of movies are very less

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.

director_name	avg_imdb	percentile	count_movie
John Landis	8.7	0.999	3
Sam Mendes	8.6	0.999	8
Wes Anderson	8.6	0.999	7
Spike Lee	8.5	0.998	15
Peter Ho-Sun Chan	8.5	0.998	1
Paul Weitz	8.5	0.998	6
Tate Taylor	8.5	0.998	2
Bruce Dellis	8.433333333	0.997	1
Martin Scorsese	8.433333333	0.997	16
Trent Cooper	8.433333333	0.997	1
Dominic Sena	8.425	0.995	3
Nick Cassavetes	8.425	0.995	4
Howard Zieff	8.425	0.995	2
Damian Nieman	8.425	0.995	1
William Cottrell	8.425	0.995	1
Alan Parker	8.425	0.995	3
Paul Greengrass	8.425	0.995	7
Morten Tyldum	8.425	0.995	2
Len Wiseman	8.4	0.994	3
Terry George	8.4	0.994	1

Report & Data Story

- ❖ There are only few number of director who has high imdb rating when compared to others
- ❖ Most of the director with high ratings have less movies only few of the directors with high rating have movies more than 10

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.

movie_title	top-10 movies
Avatar	523505847
Jurassic World	502177271
Titanic	458672302
Kinsey	449935665
Insidious:chapter 3	424449459
The Avengers	403279547
The Lion King	377783777
Star Wars: Episode I - The Phantom Menace	359544677
The Dark Knight	348316061
The Hunger Games	329999255

Report & Data Story

- ❖ Top 10 movies based on financial success are given in photo
- ❖ Avatar has the highest financial success then after avatar Jurassic world & Titanic

Results and Conclusion

- ❖ The factors that influence the success of the movies are genres,language,budget,duration,director
- ❖ Most of them follow the same trend like making more movies on drama genre only instead of genres like war,history,biography which has more imdb ratings average.
- ❖ And also there very few numbers of movies in languages other than English but their Imdb rating average is good compared to English language
- ❖ Duration of the movie also plays the role in the success
- ❖ More number of movies should be made of other genres like war, history and biography it has high average rating
- ❖ And also more number of movies should be made in other languages there is high success rate for movies other than english language.

Hyperlink For Excel Sheet

- ❖ <mailto:https://docs.google.com/spreadsheets/d/1RxkqWHmiPS9mnu0u9TgZIy5JekxMkNQP/edit?usp=drivesdk&ouid=113826139200146158008&rtpof=true&sd=true>

HyperLink For Video

- ❖ [mailto:https://drive.google.com/file/d/1S72D4WmgpCKpYmVnG1SONm4Oyb2AEQO5
/view?usp=drivesdk](mailto:https://drive.google.com/file/d/1S72D4WmgpCKpYmVnG1SONm4Oyb2AEQO5/view?usp=drivesdk)