

IMDB Movie Analysis

Project Description:

- This project aims to analyze the factors that contribute to the success of movies on IMDb, with success measured by high IMDb ratings.
- Understanding these factors is crucial for industry stakeholders such as movie producers, directors, and investors to make strategic decisions for their upcoming projects.

Approach:

I initiated the project by downloading the provided dataset and leveraging Microsoft Excel for comprehensive data analysis.

1. **Data Exploration:** Examined the data containing records of IMDB movies.
2. **Data cleaning:** Identified if there are any white spaces, duplicate or badly formatted data and cleared all those blanked data.
3. **Analyzed the data:** Used Microsoft Excel to explore the dataset, generating summary statistics, creating pivot tables and visualizing trends and patterns through charts and graphs.
4. **Insight Extraction:** Analyzed the results to identify factors such as influencing IMDb ratings and providing actionable insights for stakeholders in the movie industry.

Tech-Stack Used:

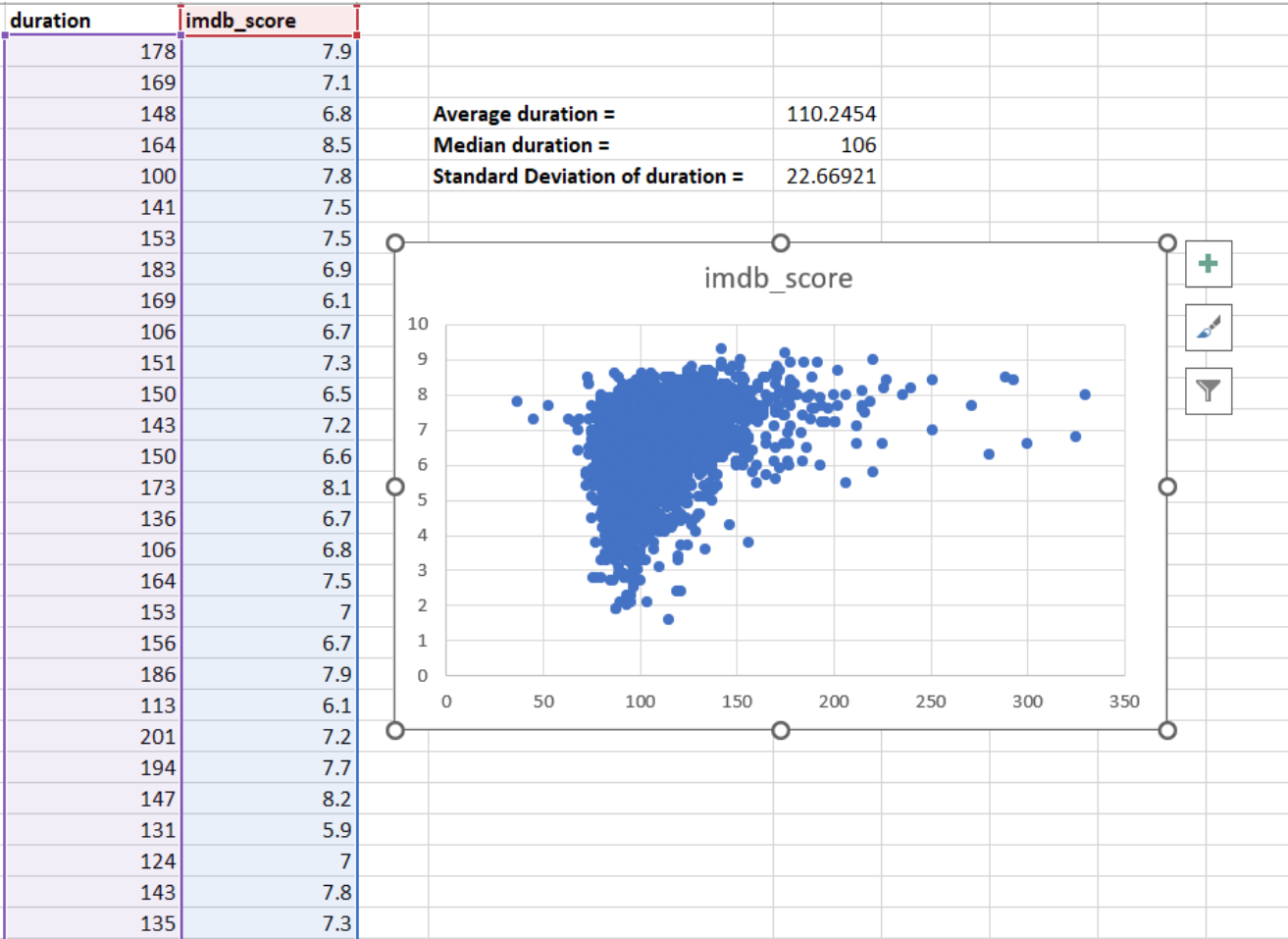
- Microsoft Excel was selected for its versatility in handling data analysis tasks.
- The software's features, including pivot tables for summarization and charts for visualization, proved instrumental in extracting meaningful insights from the hiring process dataset.

Insights:

A. **Movie Genre Analysis:** Determined the distribution of movie genres.

Genre(Separated)	Count		
Action	949		
Adventure	771		
Drama	1876		
Animation	196		
Comedy	1455	Average =	506
Mystery	378	Median =	382
Crime	704	Mode =	147
Biography	238	Max =	1876
Fantasy	504	Min =	1
Documentary	45	Var =	243320.2
Sci-Fi	491	Stdev =	493.275
Horror	386		
Romance	849		
Family	440		
Western	57		
Musical	96		
Thriller	1103		
History	147		
Music	149		
War	150		
Sport	147		
Film-Noir	1		

B. **Movie Duration Analysis:** Identified the distribution of movie duration and its impact on the IMDB score.



C. Language Analysis: Derived the distribution of movies based on their language and analyzed their impact on the Imdb score using descriptive statistics.

Language	Count of language	Median of the imdb_score	Mean of imdb_score	StdDevp of imdb_score
Aboriginal	2	6.95	6.95	0.55
Arabic	1	7.2	7.2	0
Aramaic	1	7.1	7.1	0
Bosnian	1	4.3	4.3	0
Cantonese	7	7.3	7.342857143	0.324509048
Czech	1	7.4	7.4	0
Danish	3	8.1	7.9	0.43204938
Dari	2	7.5	7.5	0.1
Dutch	3	7.8	7.566666667	0.329983165
English	3564	6.5	6.427525253	1.048039398
Filipino	1	6.7	6.7	0
French	34	7.3	7.355882353	0.51173935
German	10	7.8	7.77	0.67535176
Hebrew	1	8	8	0
Hindi	5	7.4	7.22	0.716658915
Hungarian	1	7.1	7.1	0
Indonesian	2	7.9	7.9	0.3
Italian	7	7	7.185714286	1.069617517
Japanese	10	8	7.66	0.939361485
Kazakh	1	6	6	0
Korean	5	7.7	7.7	0.509901951
Mandarin	14	7.25	7.021428571	0.737930089
Maya	1	7.8	7.8	0
Mongolian	1	7.3	7.3	0
None	1	8.5	8.5	0
Norwegian	4	7.3	7.15	0.497493719
Persian	3	8.4	8.133333333	0.449691252
Portuguese	5	8	7.76	0.875442745
Romanian	1	7.9	7.9	0

D. Director Analysis: Identified the top 10 directors based on their average IMDB score.

Directors Name	Average of imdb_score
Akira Kurosawa	8.7
Tony Kaye	8.6
Charles Chaplin	8.6
Alfred Hitchcock	8.5
Ron Fricke	8.5
Majid Majidi	8.5
Damien Chazelle	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
Richard Marquand	8.4
Asghar Farhadi	8.4

E. Budget Analysis: Calculated the profit margin of movies and identified the movie with highest profit margin and the correlation between movie budgets and gross earning.

Movie Name	Gross	Budget	Profit_margin		
Avatar	760505847	237000000	523505847		
Jurassic World	652177271	150000000	502177271		
Titanic	658672302	200000000	458672302	Movie with Highest profit_margin	
Star Wars: Episode IV - A New Hope	460935665	110000000	449935665	Avatar	523505847
E.T. the Extra-Terrestrial	434949459	105000000	424449459		
The Avengers	623279547	220000000	403279547		
The Lion King	422783777	450000000	377783777	Correlation =	0.117875345
The Jungle Book	725290282	350000000	375290282		
Star Wars: Episode I - The Phantom Menace	474544677	115000000	359544677		
The Dark Knight	533316061	185000000	348316061		
The Hunger Games	407999255	78000000	329999255		
Twilight	382898950	74000000	308898950		
Deadpool	363024263	58000000	305024263		
The Hunger Games: Catching Fire	424645577	130000000	294645577		
Jurassic Park	356784000	63000000	293784000		
Despicable Me 2	368049635	76000000	292049635		
American Sniper	350123553	58800000	291323553		
Finding Nemo	380838870	94000000	286838870		
Shrek 2	436471036	150000000	286471036		
The Lord of the Rings: The Return of the King	377019252	94000000	283019252		
Star Wars: Episode VI - Return of the Jedi	309125409	32500000	276625409		
Forrest Gump	329691196	55000000	274691196		
Star Wars: Episode V - The Empire Strikes Back	290158751	18000000	272158751		
Juno	286985680	15000000	271985680		
Alice in Wonderland	668370412	400000000	268370412		
Home Alone	285761243	18000000	267761243		
Star Wars: Episode III - Revenge of the Sith	380262555	113000000	267262555		
Spider-Man	403706375	139000000	264706375		
Minions	336029560	74000000	262029560		

Result:

Though this project was conducted entirely in Excel, gained valuable insights into the factors influencing IMDb ratings, providing actionable information for stakeholders in the movie industry to enhance the success of their future projects.

Link:

<https://docs.google.com/spreadsheets/d/1ucHFVSYXSmD-3CwQbAHE7hT9KMRGENgl/edit?usp=drivesdk&ouid=100751770545539670521&rtpof=true&sd=true>