IMDB Movie Analysis

Final Project-1

- ## Description:

  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.

- ## Approach:

  ### Data Cleaning:

  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.

  Total Rows = 5044
  Duplicate Rows = 45
  Remaining Rows = 4998
  After Remove blanks row = 3892

  ### Data Analysis:

  Here, you'll explore the data to understand the relationships between different variables. You might look at the correlation between movie ratings and other factors like genre, director, budget, etc. You might also want to consider the year of release, the actors involved, and other relevant factors.

  ### Data Visualize:

  Here , I use Pivot Chart and excel functions to visualize my data that is easily visible and understandable for anyone.

• Tech-Stack Used:

Microsoft Excel 2019

Query Editor(Inbuilt in Excel)

• Data Analytics Tasks / Insights

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.

Result:

Unique_gener	No of Movie	Avg	Median	Mode	Max	Min	Var	Std
Action	970	6.29	6.35	6.6	9	2.1	1.08	1.04
Adventure	795	6.46	6.6	6.7	8.9	2.3	1.23	1.11
Animation	199	6.70	6.8	6.7	8.6	2.8	0.97	0.99
Biography	244	7.14	7.2	7	8.9	4.5	0.50	0.71
Comedy	1511	6.18	6.3	6.7	8.8	1.9	1.08	1.04
Crime	720	6.55	6.6	6.6	9.3	2.4	0.97	0.98
Documentary	67	7.01	7.2	6.6	8.5	1.6	1.42	1.19
Drama	1961	6.78	6.9	6.7	9.3	2.1	0.81	0.90
Family	453	6.21	6.3	5.4	8.6	1.9	1.34	1.16
Fantasy	517	6.28	6.4	6.7	8.9	2.2	1.29	1.13
Film-Noir	1	7.70	7.7	"No Repeatation"	7.7	7.7	0.00	0.00
History	156	7.13	7.2	7.7	8.9	5.5	0.46	0.68
Horror	397	5.93	6	5.9	8.6	2.3	0.99	1.00
Music	161	6.38	6.5	6.5	8.5	1.6	1.45	1.20
Musical	103	6.56	6.7	7.1	8.5	2.1	1.29	1.14
Mystery	390	6.47	6.5	6.6	8.6	3.1	1.03	1.02
Romance	888	6.43	6.5	6.5	8.5	2.1	0.93	0.96
Sci-Fi	501	6.32	6.4	6.7	8.8	1.9	1.34	1.16
Short	2	6.80	6.8	"No Repeatation"	7.1	6.5	0.09	0.30
Thriller	1130	6.38	6.4	6.5	9	2.7	0.94	0.97
War	162	7.05	7.1	7.1	8.6	4.3	0.65	0.80
Western	62	6.74	6.75	6	8.9	4.1	0.94	0.97

Formula Used:

Avg =AVERAGEIF(Table4[[#All],[genres]],B4,Table4[[#All],[imdb_score]])

Median =MEDIAN(IF(B4=Table4[[#All],[genres]],Table4[[#All],[imdb_score]]))

Mode =MODE(IF(B4=Table4[[#All],[genres]],Table4[[#All],[imdb_score]]))

Max =MAX(IF(B4=Table4[[#All],[genres]],Table4[[#All],[imdb_score]]))

Min =MIN(IF(B4=Table4[[#All],[genres]],Table4[[#All],[imdb_score]]))

Var =VAR.P(IF(B4=Table4[[#All],[genres]],Table4[[#All],[imdb_score]]))

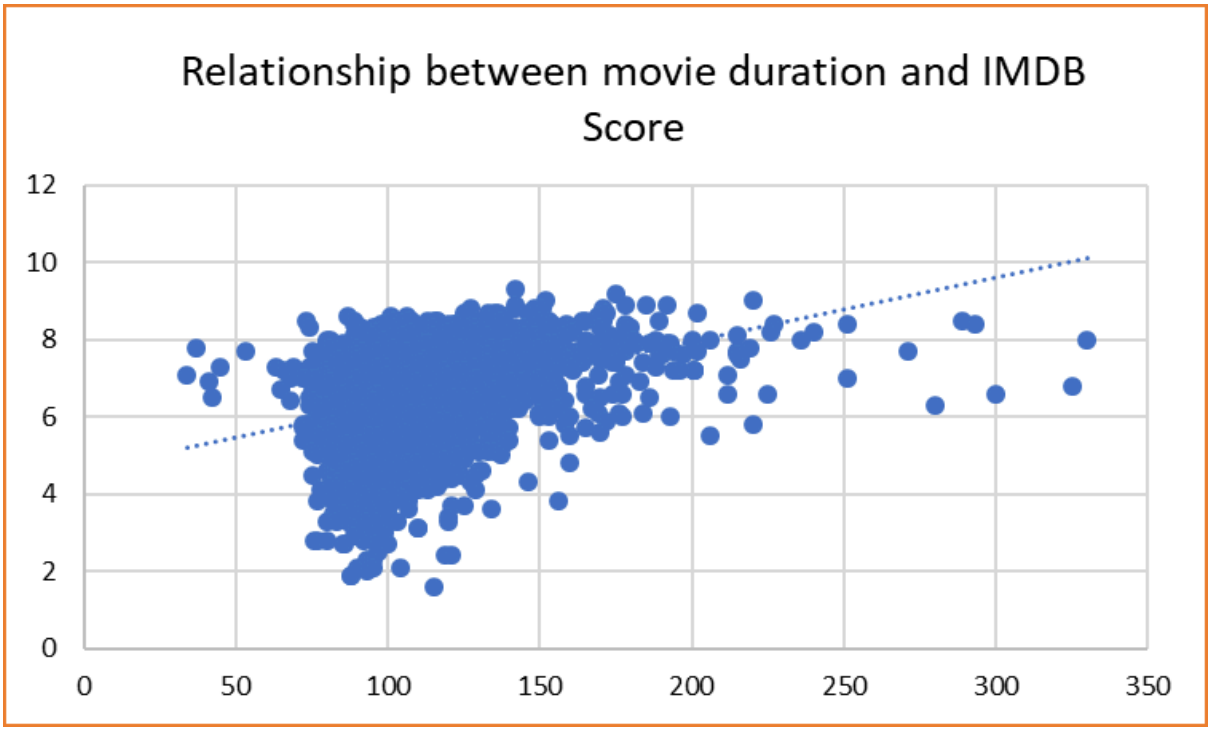
Std Dev =STDEV.P(IF(B4=Table4[[#All],[genres]],Table4[[#All],[imdb_score]]))

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.

Result:

Avg	109.8964
Median	106
Std	22.70386



Formula Used:

Avg =AVERAGE(A:A)

Median =MEDIAN(A:A)

Std Dev =STDEV.P(A:A)

C. 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.

Result:

Unique_Languaage	No of Movie	Mean	Median	Std
Aboriginal	2	6.95	6.95	0.55
Arabic	1	7.20	7.20	0.00
Aramaic	1	7.10	7.10	0.00
Bosnian	1	4.30	4.30	0.00
Cantonese	8	7.24	7.30	0.41
Czech	1	7.40	7.40	0.00
Danish	3	7.90	8.10	0.43
Dari	2	7.50	7.50	0.10
Dutch	3	7.57	7.80	0.33
Dzongkha	1	7.50	7.50	0.00
English	3707	6.42	6.50	1.05
Filipino	1	6.70	6.70	0.00
French	37	7.29	7.20	0.55
German	13	7.69	7.70	0.62
Hebrew	3	7.50	7.30	0.36
Hindi	10	6.76	7.05	1.05
Hungarian	1	7.10	7.10	0.00
Icelandic	1	6.90	6.90	0.00
Indonesian	2	7.90	7.90	0.30
Italian	7	7.19	7.00	1.07
Japanese	12	7.63	7.80	0.86
Kazakh	1	6.00	6.00	0.00
Korean	5	7.70	7.70	0.51
Mandarin	15	7.08	7.40	0.75
Maya	1	7.80	7.80	0.00
Mongolian	1	7.30	7.30	0.00
None	1	8.50	8.50	0.00
Norwegian	4	7.15	7.30	0.50
Persian	3	8.13	8.40	0.45
Portuguese	5	7.76	8.00	0.88
Romanian	1	7.90	7.90	0.00
Russian	1	6.50	6.50	0.00
Spanish	26	7.05	7.15	0.81
Swedish	1	7.60	7.60	0.00
Telugu	1	8.40	8.40	0.00
Thai	3	6.63	6.60	0.37
Vietnamese	1	7.40	7.40	0.00
Zulu	1	7.30	7.30	0.00

Formula Used:

No of Movies =COUNTIFS(K:K,A3)

Mean/Avg =AVERAGEIF(K:K,A3,L:L)

Median =MEDIAN(IF(A3=K:K,L:L))

Std Dev =STDEV.P(IF(A3=K:K,L:L))

D. 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.

Result

Here I attached the Top 20 directors name with Avg IMDB Score and on Percentile rank.

director_name	Average_IMDB_Scor	Percentile Ran
Tony Kaye	8.60	99.90%
Charles Chaplin	8.60	99.90%
Alfred Hitchcock	8.50	99.70%
Ron Fricke	8.50	99.70%
Damien Chazelle	8.50	99.70%
Majid Majidi	8.50	99.70%
Sergio Leone	8.43	99.60%
Christopher Nolan	8.43	99.60%
S.S. Rajamouli	8.40	99.30%
Richard Marquand	8.40	99.30%
Asghar Farhadi	8.40	99.30%
Marius A. Markevicius	8.40	99.30%
Lee Unkrich	8.30	99.10%
Fritz Lang	8.30	99.10%
Lenny Abrahamson	8.30	99.10%
Billy Wilder	8.30	99.10%
Pete Docter	8.23	99.00%
Hayao Miyazaki	8.23	99.00%
Quentin Tarantino	8.20	98.70%
George Roy Hill	8.20	98.70%

continue(total 1754 rows)...

Formula Used

Percentile Rank =PERCENTRANK(B:B,[@[Average_IMDB_Score]])

Avg IMDB Score=AVERAGEIF(F:F,Table13__2[@[director_name]],G:G)

E. 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	gross	budget	Profit	Correla
Avatar	760505847	237000000	523505847	0.102179
Jurassic World	652177271	150000000	502177271	
Titanic	658672302	200000000	458672302	
Star Wars: Epis	460935665	11000000	449935665	
E.T. the Extra-	434949459	10500000	424449459	
The Avengers	623279547	220000000	403279547	
The Avengers	623279547	220000000	403279547	
The Lion King	422783777	45000000	377783777	
Star Wars: Epis	474544677	115000000	359544677	
The Dark Knigh	533316061	185000000	348316061	
The Hunger Ga	407999255	78000000	329999255	
Deadpool	363024263	58000000	305024263	
The Hunger Ga	424645577	130000000	294645577	
Jurassic Park	356784000	63000000	293784000	
Despicable Me	368049635	76000000	292049635	
American Snipe	350123553	58800000	291323553	
Finding Nemo	380838870	94000000	286838870	
Shrek 2	436471036	150000000	286471036	
The Lord of the	377019252	94000000	283019252	
Star Wars: Epis	309125409	32500000	276625409	
Forrest Gump	329691196	55000000	274691196	
Star Wars: Epis	290158751	18000000	272158751	
Home Alone	285761243	18000000	267761243	
Star Wars: Epis	380262555	113000000	267262555	
Spider-Man	403706375	139000000	264706375	
Minions	336029560	74000000	262029560	
The Sixth Sense	293501675	40000000	253501675	
Jaws	260000000	8000000	252000000	
Frozen	400736600	150000000	250736600	
The Secret Life	323505540	75000000	248505540	
The Twilight Sa	296623634	50000000	246623634	
The Lord of the	340478898	94000000	246478898	

Continue(total 3892 Rows)

Top 5 Movies By Profit

Movie title	Gross	Budget	Profit
Avatar	760505847	237000000	523505847
Jurassic World	652177271	150000000	502177271
Titanic	658672302	200000000	458672302
Star Wars: Episode IV - A New Hope	460935665	11000000	449935665
E.T. the Extra-Terrestrial	434949459	10500000	424449459

AVATAR is the only movie with higher Profit...

Formula Used

Correlation =CORREL(B:B,C:C)

- ## Result

  I have done the Analysis of the given dataset on IMDB Movies data and Provide the answers and created charts as per requirement provide valuable insights to represent stakeholders and improve the data of the movies.

Drive Link: Click Here to view Excel File

THANK YOU 😀