IMDB Movie Analysis

Hyperlink of my excel sheet

Project Description:

▶ 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.

▶ Our task is to create a report that tells a story with your data. This should include initial problem, findings, and the insights I've gained. Use visualizations to help tell the story and make findings more understandable.

Tech-Stack Used:

- I've used Microsoft Power point version 2309 to create this presentation.

- I've used Microsoft Excel version 2309 to implement the task assigned.

- I chose Microsoft Excel because it is the most convenient spreadsheet and can be used efficiently to view statistics and analyze the data set given very quickly.
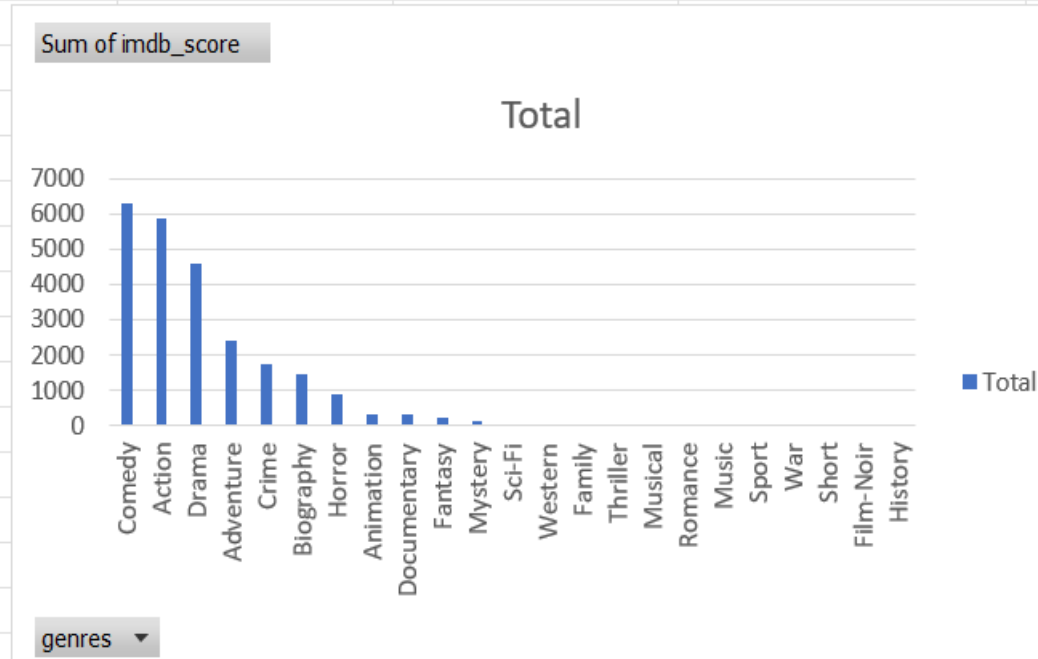
INSIGHTS

Task A: Movie Genre Analysis

- **Description:** 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.

- I've used pivot table to count the genres and then some statistical functions provided by excel to calculate the descriptive statistics.

- I've also used bar chart to show the relation between genres and their imdb scores.

- Excel Images are inserted in next slide.

A	B
Action	7.9
Action	7.1
Action	6.8
Action	8.5
Action	6.6
Action	6.2
Adventure	7.8
Action	7.5
Adventure	7.5
Action	6.9
Action	6.1
Action	6.7
Action	7.3
Action	6.5
Action	7.2
Action	6.6
Action	8.1
Action	6.7
Action	6.8
Adventure	7.5
Action	7
Action	6.7
Adventure	7.9
Adventure	6.1
Action	7.2
Drama	7.7
Action	8.2

Genres	Average of imdb_score3	Count of genres	Max of imdb_score	Min of imdb_score2	StdDevp of imdb_score2	Varp of imdb_score	Range	L	M
								Median	6.6
Drama	6.8	1911	8.8	2.1	0.9	0.8	6.9	Mode	6.7
Comedy	6.2	1492	8.8	1.9	1.0	1.1	6.3		
Thriller	5.3	1086	6.3	4.8	0.7	0.5	3.5		
Action	6.3	935	9	2.1	1.0	1.1	4.4		
Romance	6.7	866	7.1	6.2	0.4	0.2	6.9		
Adventure	6.6	766	8.6	2.3	1.1	1.3	6		
Crime	6.9	702	9.3	3.3	0.9	0.8	6.9		
Fantasy	6.2	496	7.9	4.3	0.9	0.8	6.7		
Sci-Fi	6.6	484	8.2	5	1.0	0.9	2.2		
Family	6.5	441	7.9	5.7	1.0	1.0	3.6		
Horror	5.8	379	8.5	2.3	1.0	1.0			
Mystery	6.6	376	8.5	3.3	1.1	1.2			
Biography	7.2	242	8.9	4.5	0.7	0.5	6.2		
Animation	6.8	197	8	4.5	1.0	0.9			
Music		159					0.9		
War		159					5.2		
History		152					0.9		
Sport		147					3.2		
Musical	6.8	102	7.2	6.3	0.5	0.2			
Documentary	7.0	67	8.5	1.6	1.4	2.0			
Western	6.8	58	8.9	4.1	2.0	4.0	1.5		
Short		2							
Film-Noir		1					4.8		
Grand Total	6.5	11220	9.3	1.6	1.1	1.1			

IMDB_Movies | Task A | Task B | Task C | Task D | Task E

	A	B	C	D	E	F	G	H	I
28	Action	8.2							
29	Action	5.9		Row Labels ↓	Sum of imdb_score				
30	Action	7		Comedy	6324.7				
31	Action	7.8		Action	5877.4				
32	Action	7.3		Drama	4611.5				
33	Action	7.2		Adventure	2408				
34	Adventure	6.5		Crime	1750.2				
35	Action	6.8		Biography	1473.3				
36	Adventure	7.3		Horror	906.9				
37	Action	6		Animation	311.1				
38	Action	5.7		Documentary	298.9				
39	Adventure	6.4		Fantasy	218.2				
40	Action	6.7		Mystery	145.9				
41	Action	6.8		Sci-Fi	52.7				
42	Adventure	6.3		Western	20.3				
43	Action	5.6		Family	19.5				
44	Adventure	8.3		Thriller	15.9				
45	Action	6.6		Musical	13.5				
46	Action	7.2		Romance	13.3				
47	Action	7		Music					
48	Action	8		Sport					
49	Action	7.8		War					
50	Adventure	6.3		Short					
51	Drama	7.3		Film-Noir					
52	Action	6.6		History					
53	Action	7		Grand Total	24461.3				
54	Action	6.3							

Sum of imdb_score

Total

genres ▼

IMDB_Movies | Task A | Task B | Task C | Task D | Task E | +

Task B:Movie Duration Analysis

▶ **Description :** Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

▶ To determine the relationship between movie durations and imdb score I've used multiple statistical functions like mean , median and standard deviation.

▶ I've created a scatter plot to visualize the relationship between movie duration and imdb score.

▶ Also added a trendline which direct the strength of this relationship.

▶ Image of the excel sheet is provided in next slide.

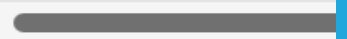|	A	B	C	D	E
	duration	imdb_score			
	178	7.9		Average of duration	109.8
	169	7.1		Median	105
	148	6.8		Std Deviation	22.8
	164	8.5			
	132	6.6			
	156	6.2			
	100	7.8			
	141	7.5			
	153	7.5			
	183	6.9			
	169	6.1			
	106	6.7			
	151	7.3			
	150	6.5			
	143	7.2			
	150	6.6			
	173	8.1			
	136	6.7			
	106	6.8			
	164	7.5			
	153	7			
	156	6.7			
	186	7.9			
	113	6.1			
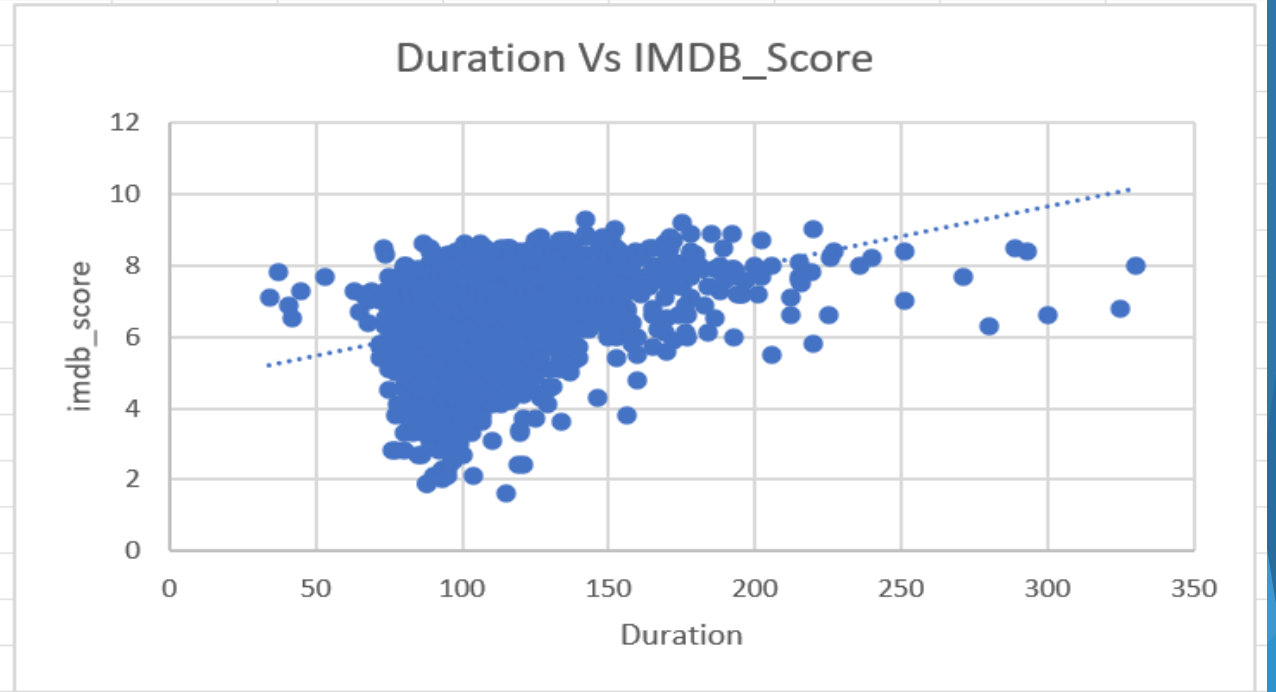
IMDB_Movies | Task A | **Task B** | Task C | Task D | Task E

Task C: Language Analysis

- **Description** : Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

- To achieve this I've used pivot table to get the count of unique languages and also to get the different statistical analysis of the data.

	language	imdb_score	movie_title	Language	Varp of imdb_score	Sum of imdb_score	StdDevp of imdb_score	Average of imdb_score	Count of movie	Median of imdb
1	language	imdb_score	movie_title	Language	Varp of imdb_score	Sum of imdb_score	StdDevp of imdb_score	Average of imdb_score	Count of movie	Median of imdb
2	English	7.9	Avatar	English	1.1	23150.1	1.1	6.4	3605	6.6
3	English	7.1	Pirates of the Caribbean: At World's End	French	0.3	269.6	0.6	7.3	37	6.6
4	English	6.8	Spectre	Spanish	0.7	183.3	0.8	7.1	26	6.6
5	English	8.5	The Dark Knight Rises	Mandarin	0.5	98.3	0.7	7.0	14	6.6
6	English	6.6	John Carter	German	0.4	100	0.6	7.7	13	6.6
7	English	6.2	Spider-Man 3	Japanese	0.7	91.5	0.9	7.6	12	6.6
8	English	7.8	Tangled	Hindi	1.1	67.6	1.1	6.8	10	6.6
9	English	7.5	Avengers: Age of Ultron	Cantonese	0.2	57.9	0.4	7.2	8	6.6
10	English	7.5	Harry Potter and the Half-Blood Prince	Italian	1.1	50.3	1.1	7.2	7	6.6
11	English	6.9	Batman v Superman: Dawn of Justice	Portuguese	0.8	38.8	0.9	7.8	5	6.6
12	English	6.1	Superman Returns	Korean	0.3	38.5	0.5	7.7	5	6.6
13	English	6.7	Quantum of Solace	Norwegian	0.2	28.6	0.5	7.2	4	6.6
14	English	7.3	Pirates of the Caribbean: Dead Man's Chest	Hebrew	0.1	22.5	0.4	7.5	3	6.6
15	English	6.5	The Lone Ranger	Persian	0.2	24.4	0.4	8.1	3	6.6
16	English	7.2	Man of Steel	Thai	0.1	19.9	0.4	6.6	3	6.6
17	English	6.6	The Chronicles of Narnia: Prince Caspian	Danish	0.2	23.7	0.4	7.9	3	6.6
18	English	8.1	The Avengers	Dutch	0.1	22.7	0.3	7.6	3	6.6
19	English	6.7	Pirates of the Caribbean: On Stranger Tides	Dari	0.0	15	0.1	7.5	2	6.6
20	English	6.8	Men in Black 3	Aboriginal	0.3	13.9	0.6	7.0	2	6.6
21	English	7.5	The Hobbit: The Battle of the Five Armies	Indonesian	0.1	15.8	0.3	7.9	2	6.6
22	English	7	The Amazing Spider-Man	Filipino	0.0	6.7	0.0	6.7	1	6.6
23	English	6.7	Robin Hood	Russian	0.0	6.5	0.0	6.5	1	6.6
24	English	7.9	The Hobbit: The Desolation of Smaug	Icelandic	0.0	6.9	0.0	6.9	1	6.6
25	English	6.1	The Golden Compass	Kazakh	0.0	6	0.0	6.0	1	6.6
26	English	7.2	King Kong	Swedish	0.0	7.6	0.0	7.6	1	6.6
27	English	7.7	Titanic	Bosnian	0.0	4.3	0.0	4.3	1	6.6
28	English	8.2	Captain America: Civil War	Zulu	0.0	7.3	0.0	7.3	1	6.6
29	English	5.9	Battleship	Arabic	0.0	7.2	0.0	7.2	1	6.6
30	English	7	Jurassic World	Romanian	0.0	7.9	0.0	7.9	1	6.6
31	English	7.8	Skyfall	Maya	0.0	7.8	0.0	7.8	1	6.6
32	English	7.3	Spider-Man 2	Czech	0.0	7.4	0.0	7.4	1	6.6
33	English	7.2	Iron Man 3	Mongolian	0.0	7.3	0.0	7.3	1	6.6
34	English	6.5	Alice in Wonderland	Telugu	0.0	8.4	0.0	8.4	1	6.6
35	English	6.8	X-Men: The Last Stand	None	0.0	8.5	0.0	8.5	1	6.6

IMDB_Movies | Task A | Task B | Task C | Task D | Task E

Task D: Director Analysis

▶ **Description** : Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

▶ I've used pivot table to get the names of top directors who have highest imdb scores.

▶ I've used percentile function to get the contribution of directors to the movies.

▶ There is also a pie chart shows the films of directors with highest imdb scores to the total no. of films.

	A	B	C	D	E	F	G	H	I
1	movie_title	director_name	imdb_score	Directors ⤓	Average of imdb_score Percentile		8.2		
2	Avatar	James Cameron	7.9	Tony Kaye	8.6			List of top directors with highest imdb_score	
3	Pirates of the Caribbean: At World's End	Gore Verbinski	7.1	Charles Chaplin	8.6			Directors ⤓	Average of imdb_score
4	Spectre	Sam Mendes	6.8	Ron Fricke	8.5			Tony Kaye	8.6
5	The Dark Knight Rises	Christopher Nolan	8.5	Majid Majidi	8.5			Charles Chaplin	8.6
6	John Carter	Andrew Stanton	6.6	Damien Chazelle	8.5			Majid Majidi	8.5
7	Spider-Man 3	Sam Raimi	6.2	Alfred Hitchcock	8.5			Ron Fricke	8.5
8	Tangled	Nathan Greno	7.8	Sergio Leone	8.4			Damien Chazelle	8.5
9	Avengers: Age of Ultron	Joss Whedon	7.5	Christopher Nolan	8.4			Alfred Hitchcock	8.5
10	Harry Potter and the Half-Blood Prince	David Yates	7.5	S.S. Rajamouli	8.4			Sergio Leone	8.4
11	Batman v Superman: Dawn of Justice	Zack Snyder	6.9	Richard Marquand	8.4			Christopher Nolan	8.4
12	Superman Returns	Bryan Singer	6.1	Marius A. Markevicius	8.4			Asghar Farhadi	8.4
13	Quantum of Solace	Marc Forster	6.7	Asghar Farhadi	8.4			S.S. Rajamouli	8.4
14	Pirates of the Caribbean: Dead Man's Chest	Gore Verbinski	7.3	Lee Unkrich	8.3			Richard Marquand	8.4
15	The Lone Ranger	Gore Verbinski	6.5	Lenny Abrahamson	8.3			Marius A. Markevicius	8.4
16	Man of Steel	Zack Snyder	7.2	Fritz Lang	8.3			Billy Wilder	8.3
17	The Chronicles of Narnia: Prince Caspian	Andrew Adamson	6.6	Billy Wilder	8.3			Lenny Abrahamson	8.3
18	The Avengers	Joss Whedon	8.1	Pete Docter	8.2			Fritz Lang	8.3
19	Pirates of the Caribbean: On Stranger Tides	Rob Marshall	6.7	Hayao Miyazaki	8.2			Lee Unkrich	8.3
20	Men in Black 3	Barry Sonnenfeld	6.8	Quentin Tarantino	8.2			Pete Docter	8.2
21	The Hobbit: The Battle of the Five Armies	Peter Jackson	7.5	Juan José Campanella	8.2			Hayao Miyazaki	8.2
22	The Amazing Spider-Man	Marc Webb	7	Joshua Oppenheimer	8.2			Grand Total	8.4
23	Robin Hood	Ridley Scott	6.7	George Roy Hill	8.2				
24	The Hobbit: The Desolation of Smaug	Peter Jackson	7.9	Elia Kazan	8.2				
25	The Golden Compass	Chris Weitz	6.1	Victor Fleming	8.2			Total no. of films	3785
26	King Kong	Peter Jackson	7.2	Milos Forman	8.1			No. of highly successful films	88
27	Titanic	James Cameron	7.7	William Wyler	8.1				
28	Captain America: Civil War	Anthony Russo	8.2	Tim Miller	8.1				
29	Battleship	Peter Berg	5.9	Terry George	8.1				
30	Jurassic World	Colin Trevorrow	7	Michael Wadleigh	8.1				
31	Skyfall	Sam Mendes	7.8	Michael Roemer	8.1				
32	Spider-Man 2	Sam Raimi	7.3	Je-kyu Kang	8.1				
33	Iron Man 3	Shane Black	7.2	David Sington	8.1				
34	Alice in Wonderland	Tim Burton	6.5	Akira Kurosawa	8.1				
35	X-Men: The Last Stand	Brett Ratner	6.8	Stephen Chbosky	8.0				
36	Monsters University	Dan Scanlon	7.3	Vincent Paronnaud	8.0				
37	Transformers: Revenge of the Fallen	Michael Bay	6	Michel Hazanavicius	8.0				
38	Transformers: Age of Extinction	Michael Bay	5.7	David Lean	8.0				
39	Oz the Great and Powerful	Sam Raimi	6.4	Ari Folman	8.0				
40	The Amazing Spider-Man 2	Marc Webb	6.7	Frank Darabont	8.0				
41	TRON: Legacy	Joseph Kosinski	6.8	Denis Villeneuve	8.0				

Chart Title

- Total no. of films - No. of highly successful films

IMDB_Movies   Task A   Task B   Task C   **Task D**   Task E

Task E:Budget Analysis

- **Description:** Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

- I've used CORREL function given in excel to calculate the correlation between budget and gross earnings.

- I've also calculated profit of all the movies by subtracting the gross earning with budgets .

- The movie with the highest profit is AVATAR.

- The data is sorted in descending order on the basis of profit margin.

	A	B	C	D	E	F	G
1	movie_title	budget	gross earning	Profit_Margin			
2	Avatar	237000000	760505847	523505847	Correlation between budget and gross	0.100942	
3	Jurassic World	150000000	652177271	502177271	Movie with highest profit margin	5.24E+08	Avatar
4	Titanic	200000000	658672302	458672302			
5	Star Wars: Episode IV - A New Hope	11000000	460935665	449935665			
6	E.T. the Extra-Terrestrial	10500000	434949459	424449459			
7	The Avengers	220000000	623279547	403279547			
8	The Lion King	45000000	422783777	377783777			
9	Star Wars: Episode I - The Phantom Menace	115000000	474544677	359544677			
10	The Dark Knight	185000000	533316061	348316061			
11	The Hunger Games	78000000	407999255	329999255			
12	Deadpool	58000000	363024263	305024263			
13	The Hunger Games: Catching Fire	130000000	424645577	294645577			
14	Jurassic Park	63000000	356784000	293784000			
15	Despicable Me 2	76000000	368049635	292049635			
16	American Sniper	58800000	350123553	291323553			
17	Finding Nemo	94000000	380838870	286838870			
18	Shrek 2	150000000	436471036	286471036			
19	The Lord of the Rings: The Return of the King	94000000	377019252	283019252			
20	Star Wars: Episode VI - Return of the Jedi	32500000	309125409	276625409			
21	Forrest Gump	55000000	329691196	274691196			
22	Star Wars: Episode V - The Empire Strikes Back	18000000	290158751	272158751			
23	Home Alone	18000000	285761243	267761243			
24	Star Wars: Episode III - Revenge of the Sith	113000000	380262555	267262555			
25	Spider-Man	139000000	403706375	264706375			
26	Minions	74000000	336029560	262029560			
27	The Sixth Sense	40000000	293501675	253501675			
28	Jaws	8000000	260000000	252000000			
29	Frozen	150000000	400736600	250736600			

Results:

▶ This project has many conclusion drawn out of it like profits, most common language movies viewed all over world, impact of director on the imdb score of movie, stats about the genres etc.

▶ This project made me experience a real time data analysis of a thousand of movie and filter them on the basis of their properties which will be further used by the marketing team to push them to the audience.

THANK YOU

-BY AVANI SHARMA

B.TECH in Computer Science And Engineering
My Email
Contact- 8435725000