IMDB MOVIE ANALYSIS

Project Description:
In this project, we are supposed to make a detailed report of movies which helps to answer very important questions like which movie has made huge profit, who is the best director, what are the most famous genres etc.

A. **Cleaning the data:: PThis is one of the most important steps to perform before moving forward with the analysis. Use your knowledge learned till now to do this. (Dropping columns, removing null values, etc.)**
Your task: Clean the data

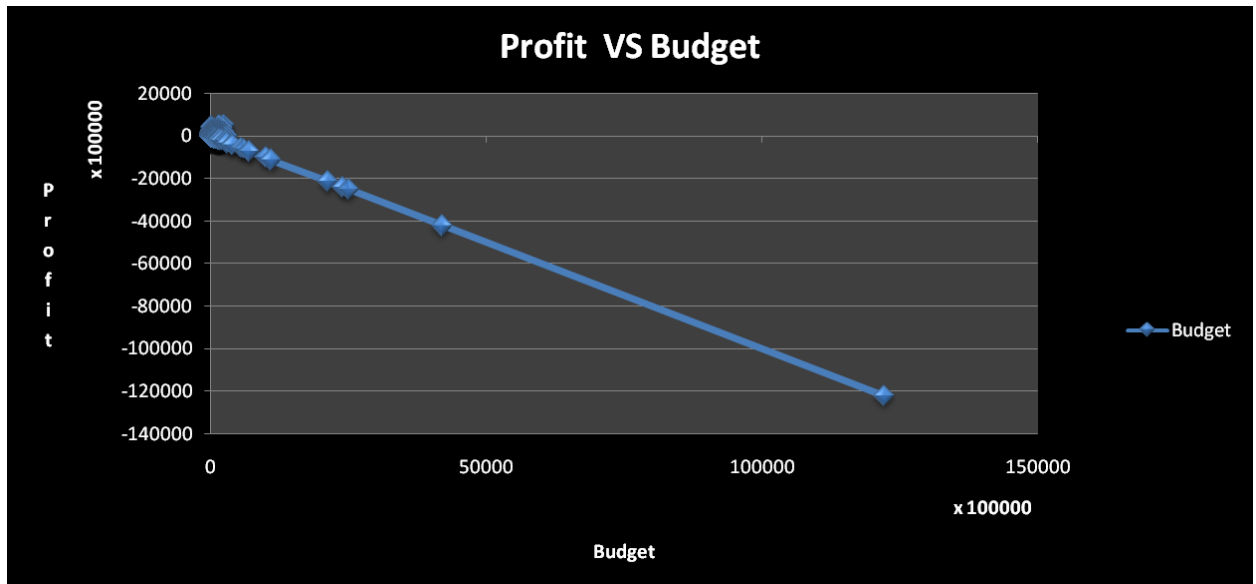|	actor_1_name	movie_title	num_voted_users	num_user_for_review	language	country	budget	title_year	imdb_score
2	CCH Pounder	Avatar	886204	3054	English	USA	237000000	2009	7.9
3	Bryce Dallas Howard	Jurassic World	418214	1290	English	USA	150000000	2015	7
4	Leonardo DiCaprio	Titanic	793059	2528	English	USA	200000000	1997	7.7
5	Harrison Ford	Star Wars: Episode IV - A New Hope	911097	1470	English	USA	11000000	1977	8.7
6	Henry Thomas	E.T. the Extra-Terrestrial	281842	515	English	USA	10500000	1982	7.9
7	Chris Hemsworth	The Avengers	995415	1722	English	USA	220000000	2012	8.1
8	Matthew Broderick	The Lion King	644348	656	English	USA	45000000	1994	8.5
9	Natalie Portman	Star Wars: Episode I - The Phantom Menace	534658	3597	English	USA	115000000	1999	6.5
10	Christian Bale	The Dark Knight	1676169	4667	English	USA	185000000	2008	9
11	Jennifer Lawrence	The Hunger Games	701607	1959	English	USA	78000000	2012	7.3
12	Ryan Reynolds	Deadpool	479047	1058	English	USA	58000000	2016	8.1
13	Jennifer Lawrence	The Hunger Games: Catching Fire	498397	706	English	USA	130000000	2013	7.6
14	Wayne Knight	Jurassic Park	613473	895	English	USA	63000000	1993	8.1
15	Steve Carell	Despicable Me 2	286877	284	English	USA	76000000	2013	7.5
16	Bradley Cooper	American Sniper	325264	916	English	USA	58800000	2014	7.3
17	Alexander Gould	Finding Nemo	692482	866	English	USA	94000000	2003	8.2
18	Rupert Everett	Shrek 2	314630	483	English	USA	150000000	2004	7.2
19	Orlando Bloom	The Lord of the Rings: The Return of the King	1215718	3189	English	USA	94000000	2003	8.9
20	Harrison Ford	Star Wars: Episode VI - Return of the Jedi	681857	647	English	USA	32500000	1983	8.4
21	Tom Hanks	Forrest Gump	1251222	1398	English	USA	55000000	1994	8.8
22	Harrison Ford	Star Wars: Episode V - The Empire Strikes Back	837759	900	English	USA	18000000	1980	8.8
23	Macaulay Culkin	Home Alone	311075	305	English	USA	18000000	1990	7.5
24	Natalie Portman	Star Wars: Episode III - Revenge of the Sith	520104	3286	English	USA	113000000	2005	7.6
25	J.K. Simmons	Spider-Man	544665	2012	English	USA	139000000	2002	7.3
26	Steve Carell	Minions	142403	275	English	USA	74000000	2015	6.4
27	Bruce Willis	The Sixth Sense	704766	2073	English	USA	40000000	1999	8.1
28	Roy Scheider	Jaws	412454	962	English	USA	8000000	1975	8
29	Josh Gad	Frozen	421658	904	English	USA	150000000	2013	7.6

I have dropped some useless columns and also removed null values.

B. **Movies with highest profit:** Create a new column called profit which contains the difference of the two columns: gross and budget. Sort the column using the profit column as reference. Plot profit (y-axis) vs budget (x- axis) and observe the outliers using the appropriate chart type.
Your task: Find the movies with the highest profit?

Find the movies with the highest profit?

Avatar

Profit VS Budget

C. **Top 250:** Create a new column IMDb_Top_250 and store the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb_score). Also make sure that for all of these movies, the num_voted_users is greater than 25,000. Also add a Rank column containing the values 1 to 250 indicating the ranks of the corresponding films.

Extract all the movies in the IMDb_Top_250 column which are not in the English language and store them in a new column named Top_Foreign_Lang_Film. You can use your own imagination also!
Your task: Find IMDB Top 250

IMDb_Top_250	Rank		Top_Foreign_Lang_Film
The Shawshank Redemption	1		The Good, the Bad and the Ugly
The Godfather	2		City of God
The Dark Knight	3		Seven Samurai
The Godfather: Part II	3		Spirited Away
Pulp Fiction	5		The Lives of Others
The Lord of the Rings: The Return of the King	5		Children of Heaven
Schindler's List	5		AmÃ©lie
The Good, the Bad and the Ugly	5		Oldboy
Inception	9		Princess Mononoke
Fight Club	9		Das Boot
Forrest Gump	9		A Separation
The Lord of the Rings: The Fellowship of the Rir	9		Baahubali: The Beginning
Star Wars: Episode V - The Empire Strikes Back	9		Downfall
The Matrix	14		The Hunt
The Lord of the Rings: The Two Towers	14		Metropolis
Star Wars: Episode IV - A New Hope	14		Pan's Labyrinth
Goodfellas	14		Howl's Moving Castle
One Flew Over the Cuckoo's Nest	14		The Secret in Their Eyes
City of God	14		Incendies
Seven Samurai	14		Amores Perros
Se7en	21		Akira
Interstellar	21		Elite Squad
The Silence of the Lambs	21		The Celebration
Saving Private Ryan	21		The Sea Inside
American History X	21		Tae Guk Gi: The Brotherhood of V
The Usual Suspects	21		A Fistful of Dollars
Spirited Away	21		Persepolis
Modern Times	21		My Name Is Khan
The Dark Knight Rises	29		Waltz with Bashir
Gladiator	29		Central Station
Django Unchained	29		Crouching Tiger, Hidden Dragon
The Departed	29		Hero

D. **Best Directors:** TGroup the column using the director_name column.

Find out the top 10 directors for whom the mean of imdb_score is the highest and store them in a new column top10director. In case of a tie in IMDb score between two directors, sort them alphabetically.
Your task: Find the best directors

Top 10 Director	Mean of IMDb Score
Charles Chaplin	8.6
Tony Kaye	8.6
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Ron Fricke	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
Asghar Farhadi	8.4
Marius A. Markevicius	8.4
Richard Marquand	8.4
S.S. Rajamouli	8.4

E. Popular Genres: Perform this step using the knowledge gained while performing previous steps.
Your task: Find popular genres

Genres	Mean of IMDb Score
Adventure|Animation	8.5
Crime|Drama|Fantasy	8.5
Action|Adventure|Dra	8.4
Adventure|Animation	8.4
Adventure|Drama|Thr	8.4
Adventure|Animation	8.3
Biography|Drama|Hist	8.3
Documentary|Drama|$	8.3
Documentary|War	8.3
Adventure|Drama|Wa	8.25

F. Charts: Create three new columns namely, Meryl_Streep, Leo_Caprio, and Brad_Pitt which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors. Use only the actor_1_name column for extraction. Also, make sure that you use the names

'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' for the said extraction.

Append the rows of all these columns and store them in a new column named Combined.

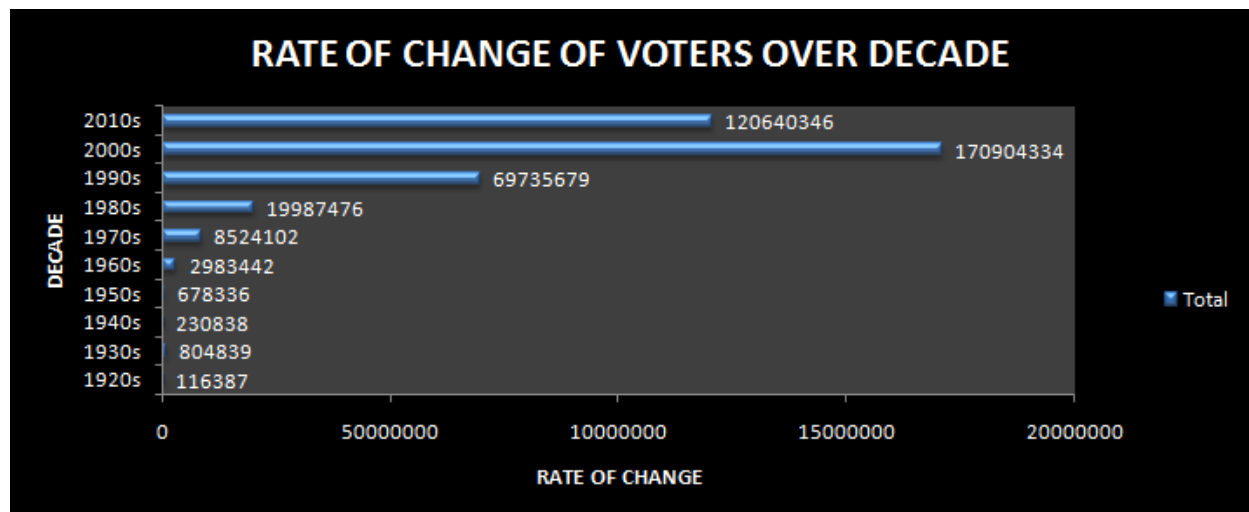Group the combined column using the actor_1_name column.

Find the mean of the num_critic_for_reviews and num_users_for_review and identify the actors which have the highest mean.

Observe the change in number of voted users over decades using a bar chart. Create a column called decade which represents the decade to which every movie belongs to. For example, the title_year year 1923, 1925 should be stored as 1920s. Sort the column based on the column decade, group it by decade and find the sum of users voted in each decade. Store this in a new data frame called df_by_decade.

Your task: Find the critic-favorite and audience-favorite actors

title_year	imdb_score	Meryl Streep	Leonardo DiCaprio	Brad Pitt	Combined	Grouping
2009	7.9	The Devil Wears Pr	Titanic	Ocean's Eleven	The Devil Wears Pr	Meryl Streep
2015	7	Out of Africa	Inception	Mr. & Mrs. Smith	Out of Africa	Meryl Streep
1997	7.7	Julie & Julia	Catch Me If You Can	Interview with the Vampir	Julie & Julia	Meryl Streep
1977	8.7	Hope Springs	Django Unchained	Fury	Hope Springs	Meryl Streep
1982	7.9	It's Complicated	The Revenant	Ocean's Twelve	It's Complicated	Meryl Streep
2012	8.1	The Iron Lady	Shutter Island	Babel	The Iron Lady	Meryl Streep
1994	8.5	The Hours	The Departed	Killing Them Softly	The Hours	Meryl Streep
1999	6.5	A Prairie Home Con	The Great Gatsby	True Romance	A Prairie Home Con	Meryl Streep
2008	9	The River Wild	The Great Gatsby	By the Sea	The River Wild	Meryl Streep
2012	7.3	One True Thing	Romeo + Juliet	The Tree of Life	One True Thing	Meryl Streep
2016	8.1	Lions for Lambs	The Man in the Iron M	The Curious Case of Benja	Lions for Lambs	Meryl Streep
2013	7.6		The Wolf of Wall Stree	Fight Club	Titanic	Leonardo DiCaprio
1993	8.1		J. Edgar	The Assassination of Jesse	Inception	Leonardo DiCaprio
2013	7.5		The Aviator	Seven Years in Tibet	Catch Me If You Car	Leonardo DiCaprio
2014	7.3		Marvin's Room	Sinbad: Legend of the Sev	Django Unchained	Leonardo DiCaprio
2003	8.2		The Beach	Troy	The Revenant	Leonardo DiCaprio
2004	7.2		Revolutionary Road	Spy Game	Shutter Island	Leonardo DiCaprio
2003	8.9		The Quick and the Dead		The Departed	Leonardo DiCaprio
1983	8.4		Gangs of New York		The Great Gatsby	Leonardo DiCaprio
1994	8.8		Body of Lies		The Great Gatsby	Leonardo DiCaprio
1980	8.8		Blood Diamond		Romeo + Juliet	Leonardo DiCaprio
1990	7.5				The Man in the Iror	Leonardo DiCaprio
2005	7.6				The Wolf of Wall St	Leonardo DiCaprio
2002	7.3				J. Edgar	Leonardo DiCaprio
2015	6.4				The Aviator	Leonardo DiCaprio
1999	8.1				Marvin's Room	Leonardo DiCaprio
1975	8				The Beach	Leonardo DiCaprio
2013	7.6				Revolutionary Road	Leonardo DiCaprio

Actor Name	Average of num_critic_for_reviews
Albert Finney	750

Actor Name	Average of num_user_for_reviews
Heather Donahue	3400

RATE OF CHANGE OF VOTERS OVER DECADE

DECADE	Total
2010s	120640346
2000s	170904334
1990s	69735679
1980s	19987476
1970s	8524102
1960s	2983442
1950s	678336
1940s	230838
1930s	804839
1920s	116387

RATE OF CHANGE

Find the critic favourite actor

Actor Name	Sum of num_critic_for_reviews
Johnny Depp	9555

Find the audience favourite actor

Actor Name	Sum of num_voted_users
Leonardo DiCaprio	9281136

Approach:
For this project I have used MS Excel to clean the dataset (Removing Columns and Null Values). I have used sorting, different formulae, pivots and charts to get the desired insights from the data.

Tech-Stack Used:
I have used MS Excel to perform the whole analysis using various pivots and charts.
I have also used a snipping tool to get the screenshot of the output and also used MS Word to make the whole project more attractive.

Insights and Results:
While doing this project I have learnt cleaning of data (Removing columns and null values), different formulae like if statement, average etc. I have also learnt how to make pivot tables and create beautiful charts using MS Excel.

While doing this project I have also learnt how to find best director, movie or genres are found, how to get the report of profit of different movies and how to decide the best movie according to the IMDb score.

Drive Link:
https://docs.google.com/spreadsheets/d/1TUBP9WjPaeJF9Ug79BP36YJYK-d3J4EU/edit?usp=sharing&ouid=113747271976476220784&rtpof=true&sd=true