NAME: ARIN KUMAR **DATE: 1/3/2023**

PROJECT-5

IMDB MOVIE ANALYSIS

Project Description

The project is about analyzing a dataset of movies in order to extract valuable insights and answer some business questions. The dataset includes information about movie titles, their cast, directors, ratings, budgets, and revenues. The analysis will be performed using Excel or Google Sheets.

Approach

The first step is to clean the dataset by removing null values and dropping unnecessary columns. Then, we will create a new column called "profit" to calculate the difference between gross and budget for each movie. We will use this column to find the movies with the highest profit and plot a scatter chart to observe the outliers.

Next, we will create a new column called "IMDb_Top_250" to store the top 250 movies with the highest IMDb rating and num_voted_users greater than 25,000. We will also add a new column called "Rank" to indicate the rank of each movie. We will extract all the non-English movies from this column and store them in a new column called "Top_Foreign_Lang_Film".

Then, we will group the dataset by director name and find the top 10 directors with the highest mean IMDb score. In case of a tie, we will sort them alphabetically. We will also find the popular genres by analyzing the frequency of each genre in the dataset.

After that, we will create three new columns to store the movies with the lead actors "Meryl Streep", "Leonardo DiCaprio", and "Brad Pitt". We will append

these columns and group the resulting column by actor name. Then, we will find the actors with the highest mean num_critic_for_reviews and num_users_for_review.

Finally, we will create a new column called "decade" to represent the decade to which each movie belongs. We will observe the change in the number of voted users over decades using a bar chart and find the sum of users voted in each decade. We will store this information in a new data frame called "df_by_decade".

Tech-Stack Used

The analysis was performed using Excel, which is a spreadsheet software that allows us to perform data analysis, create charts and pivot tables, and use various functions and formulas. It was chosen because it is a widely used tool that is accessible and easy to learn.

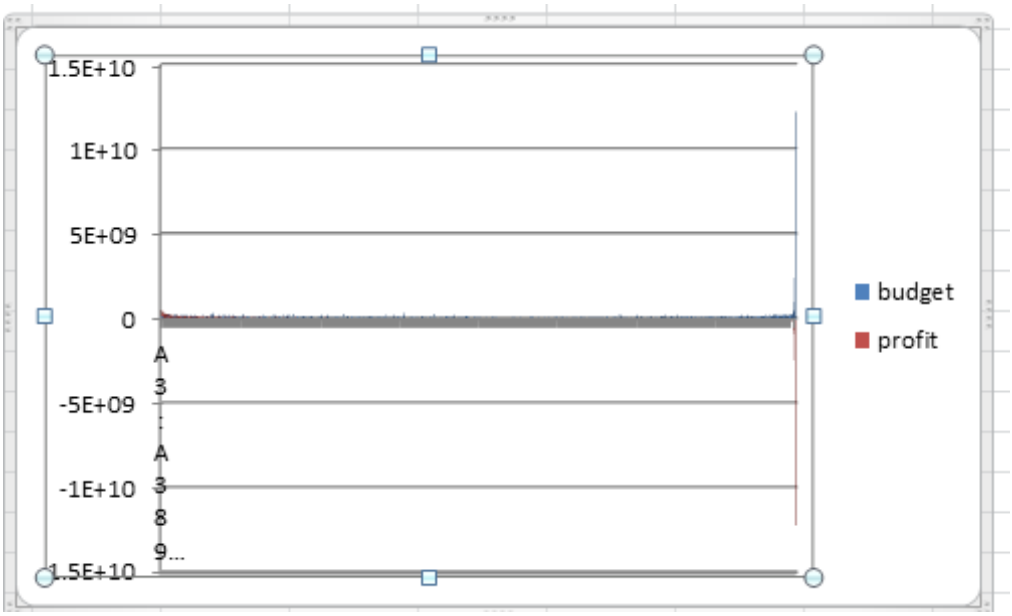
Insights

The analysis revealed several insights about the movie industry. We found that the highest profit movies have a budget of around 200 million dollars and a profit of over 1 billion dollars. We also observed some outliers with a very high profit and low budget.

A. **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?

gross	budget	profit	movie_title
7.61E+08	2.37E+08	523505847	AvatarÂ
6.52E+08	1.5E+08	502177271	Jurassic WorldÂ
6.59E+08	2E+08	458672302	TitanicÂ
4.61E+08	11000000	449935665	Star Wars: Episode IV - A New HopeÂ
4.35E+08	10500000	424449459	E.T. the Extra-TerrestrialÂ
6.23E+08	2.2E+08	403279547	The AvengersÂ
6.23E+08	2.2E+08	403279547	The AvengersÂ
4.23E+08	45000000	377783777	The Lion KingÂ
4.75E+08	1.15E+08	359544677	Star Wars: Episode I - The Phantom MenaceÂ
5.33E+08	1.85E+08	348316061	The Dark KnightÂ
4.08E+08	78000000	329999255	The Hunger GamesÂ
3.63E+08	58000000	305024263	DeadpoolÂ
4.25E+08	1.3E+08	294645577	The Hunger Games: Catching FireÂ
3.57E+08	63000000	293784000	Jurassic ParkÂ
3.68E+08	76000000	292049635	Despicable Me 2Â
3.5E+08	58800000	291323553	American SniperÂ
3.81E+08	94000000	286838870	Finding NemoÂ
4.36E+08	1.5E+08	286471036	Shrek 2Â
3.77E+08	94000000	283019252	The Lord of the Rings: The Return of the King
3.09E+08	32500000	276625409	Star Wars: Episode VI - Return of the JediÂ
3.3E+08	55000000	274691196	Forrest GumpÂ
2.9E+08	18000000	272158751	Star Wars: Episode V - The Empire Strikes Ba
2.86E+08	18000000	267761243	Home AloneÂ
3.8E+08	1.13E+08	267262555	Star Wars: Episode III - Revenge of the SithÂ



Avatar is the most profitable movie with profit of 523505847.

B. **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.

Rank	IMDB_TOP_250	num_voted_users	imdb_score
1	The Shawshank RedemptionÂ	1689764	9.3
2	The GodfatherÂ	1155770	9.2
3	The Dark KnightÂ	1676169	9
4	The Godfather: Part IIÂ	790926	9
5	FargoÂ	170055	9
6	The Lord of the Rings: The Return of the KingÂ	1215718	8.9
7	Schindler's ListÂ	865020	8.9
8	Pulp FictionÂ	1324680	8.9
9	The Good, the Bad and the UglyÂ	503509	8.9
10	12 Angry MenÂ	447785	8.9
11	InceptionÂ	1468200	8.8
12	The Lord of the Rings: The Fellowship of the RingÂ	1238746	8.8
13	DaredevilÂ	213483	8.8
14	Fight ClubÂ	1347461	8.8
15	Forrest GumpÂ	1251222	8.8
16	It's Always Sunny in PhiladelphiaÂ	133415	8.8
17	Star Wars: Episode V - The Empire Strikes BackÂ	837759	8.8
18	The Lord of the Rings: The Two TowersÂ	1100446	8.7
19	The MatrixÂ	1217752	8.7
20	Friday Night LightsÂ	42746	8.7
21	GoodfellasÂ	728685	8.7
22	Star Wars: Episode IV - A New HopeÂ	911097	8.7
23	One Flew Over the Cuckoo's NestÂ	680041	8.7
24	City of GodÂ	533200	8.7

The Shawshank Redemption is highest IMDB rated movie with 9.3.

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

TOP_FOREIGN_LANG_FILM	num_voted_users	imdb_score	language
NightcrawlerÂ	293304	7.9	Mandarin
The HangoverÂ	583341	7.8	Aboriginal
Fear and Loathing in Las VegasÂ	213226	7.7	Spanish
The NegotiatorÂ	107227	7.3	French
Bridge to TerabithiaÂ	110390	7.2	Russian
TimecrimesÂ	40878	7.2	Mandarin
We Were SoldiersÂ	103241	7.1	Mandarin
Two LoversÂ	29613	7.1	Maya
Legend of the Guardians: The Owls of Ga'HooleÂ	65785	7	French
The Prince of EgyptÂ	91093	7	Telugu
Non-StopÂ	200647	7	Mandarin
RadioÂ	32370	6.9	Spanish
Four BrothersÂ	109894	6.9	Japanese
The Best of MeÂ	43084	6.7	Aramaic
Friends with BenefitsÂ	270228	6.6	Japanese
Kiss of the DragonÂ	53126	6.6	French
Jackass: The MovieÂ	67992	6.6	Dutch
Step UpÂ	90938	6.5	Cantonese
In the Land of WomenÂ	27689	6.5	Dari
The Perfect StormÂ	133076	6.4	Japanese
ClickÂ	246492	6.4	Mandarin
Charlotte's WebÂ	27838	6.4	German
Red DawnÂ	41776	6.4	Japanese
The LosersÂ	74691	6.4	Mongolian

Nightcrawler is the highest IMDB rated non English language film with imdb rating 7.9.

C. **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_DIRECTORS	MEAN_IMDB_SCORE
Doug Walker	9.1
James Cameroon	9.1
Gore Verbinski	9
Nathan Greno	9
Sam Mendes	8.95
Joss Whedon	8.9
Andrew Stanton	8.8
Rob Marshall	8.8
Peter Jackson	8.8
Barry Sonnenfeld	8.8

Doug Walker and James Cameroon has the joint highest mean imdb score of 9.1.

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

Row Labels	Count of genres
Action|Crime|Drama|Thriller	68
Action|Crime|Thriller	65
Comedy	209
Comedy|Drama	191
Comedy|Drama|Romance	187
Comedy|Romance	158
Crime|Drama|Thriller	101
Drama	236
Drama|Romance	152
Horror	71
Grand Total	**1438**

Drama has the highest count of genres with 236 movies.

E. **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.

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

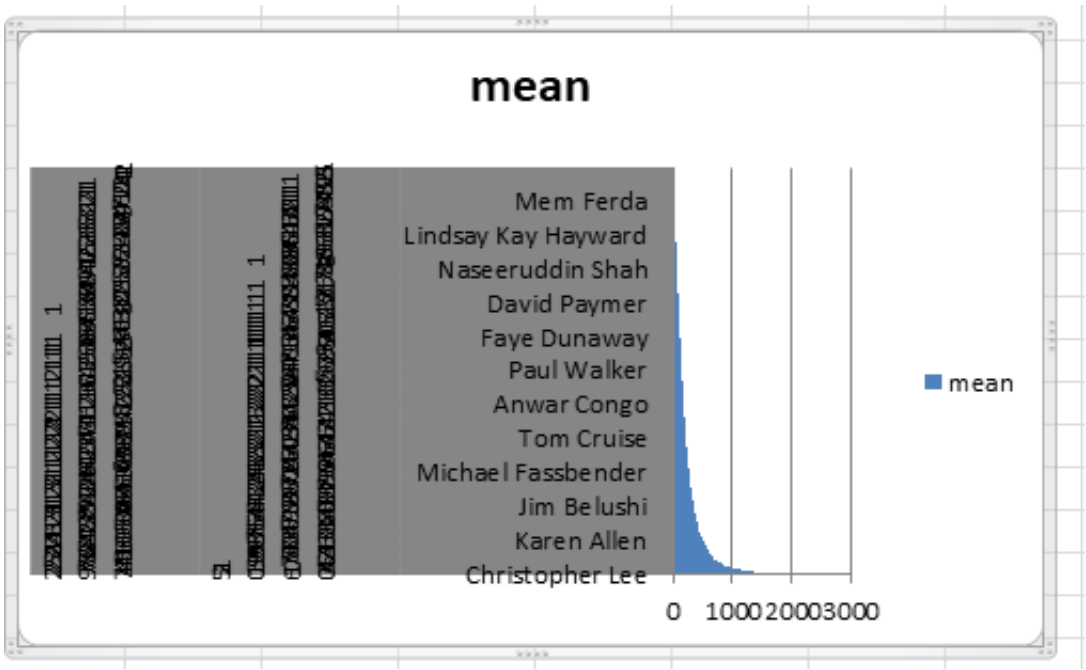
2	Meryl_Streep	Leo_Caprio	Brad_Pitt
3	It's ComplicatedÂ	TitanicÂ	The Curious Case of Benjamin ButtonÂ
4	The River WildÂ	The Great GatsbyÂ	TroyÂ
5	Julie & JuliaÂ	InceptionÂ	Ocean's TwelveÂ
6	The Devil Wears PradaÂ	The RevenantÂ	Mr. & Mrs. SmithÂ
7	Lions for LambsÂ	The AviatorÂ	Spy GameÂ
8	Out of AfricaÂ	Django UnchainedÂ	Ocean's ElevenÂ
9	Hope SpringsÂ	Blood DiamondÂ	FuryÂ
10	One True ThingÂ	The Wolf of Wall StreetÂ	Seven Years in TibetÂ
11	Florence Foster JenkinsÂ	Gangs of New YorkÂ	Fight ClubÂ
12	The HoursÂ	The DepartedÂ	Sinbad: Legend of the Seven SeasÂ
13	The Iron LadyÂ	Shutter IslandÂ	Interview with the Vampire: The Vampire ChroniclesÂ
14	A Prairie Home CompanionÂ	Body of LiesÂ	The Tree of LifeÂ
15	JuliaÂ	Catch Me If You CanÂ	The Assassination of Jesse James by the Coward Robert FordÂ
16		The BeachÂ	BabelÂ
17		Revolutionary RoadÂ	By the SeaÂ
18		The Man in the Iron MaskÂ	Killing Them SoftlyÂ
19		J. EdgarÂ	True RomanceÂ
20		The Quick and the DeadÂ	Johnny SuedeÂ
21		Marvin's RoomÂ	
22		Romeo + JulietÂ	
23		The Great GatsbyÂ	

Combined	actor_1_name
TitanicÂ	Leonardo DiCaprio
The Great GatsbyÂ	Leonardo DiCaprio
InceptionÂ	Leonardo DiCaprio
The Curious Case of Benjamin B	Brad Pitt
TroyÂ	Brad Pitt
The RevenantÂ	Leonardo DiCaprio
Ocean's TwelveÂ	Brad Pitt
Mr. & Mrs. SmithÂ	Brad Pitt
The AviatorÂ	Leonardo DiCaprio
Django UnchainedÂ	Leonardo DiCaprio
Blood DiamondÂ	Leonardo DiCaprio
The Wolf of Wall StreetÂ	Leonardo DiCaprio
Gangs of New YorkÂ	Leonardo DiCaprio
The DepartedÂ	Leonardo DiCaprio
Spy GameÂ	Brad Pitt
Ocean's ElevenÂ	Brad Pitt
It's ComplicatedÂ	Meryl Streep
Shutter IslandÂ	Leonardo DiCaprio
FuryÂ	Brad Pitt
Seven Years in TibetÂ	Brad Pitt
Body of LiesÂ	Leonardo DiCaprio
Fight ClubÂ	Brad Pitt
Sinbad: Legend of the Seven Se	Brad Pitt
Catch Me If You CanÂ	Leonardo DiCaprio

Interview with the Vampire: Th	Brad Pitt
The BeachÂ	Leonardo DiCaprio
The River WildÂ	Meryl Streep
Revolutionary RoadÂ	Leonardo DiCaprio
Julie & JuliaÂ	Meryl Streep
The Devil Wears PradaÂ	Meryl Streep
The Man in the Iron MaskÂ	Leonardo DiCaprio
J. EdgarÂ	Leonardo DiCaprio
Lions for LambsÂ	Meryl Streep
The Tree of LifeÂ	Brad Pitt
The Quick and the DeadÂ	Leonardo DiCaprio
Out of AfricaÂ	Meryl Streep
Hope SpringsÂ	Meryl Streep
One True ThingÂ	Meryl Streep
The Assassination of Jesse Jame	Brad Pitt
Florence Foster JenkinsÂ	Meryl Streep
The HoursÂ	Meryl Streep
Marvin's RoomÂ	Leonardo DiCaprio
BabelÂ	Brad Pitt
By the SeaÂ	Brad Pitt
Killing Them SoftlyÂ	Brad Pitt
Romeo + JulietÂ	Leonardo DiCaprio
The Iron LadyÂ	Meryl Streep
True RomanceÂ	Brad Pitt
A Prairie Home CompanionÂ	Meryl Streep

The Great GatsbyÂ	Leonardo DiCaprio	
JuliaÂ	Meryl Streep	
Johnny SuedeÂ	Brad Pitt	E

num_critic_for_reviews	num_voted_users	actor_name	mean
297	5060	Christopher Lee	2678.5
645	4667	Christian Bale	2656
199	4144	Morgan Freeman	2171.5
313	3646	Keanu Reeves	1979.5
320	3597	Natalie Portman	1958.5
284	3516	Natalie Portman	1900
723	3054	CCH Pounder	1888.5
360	3400	Heather Donahue	1880
673	3018	Henry Cavill	1845.5
359	3286	Natalie Portman	1822.5
328	3189	Orlando Bloom	1758.5
813	2701	Tom Hardy	1757
642	2803	Leonardo DiCaprio	1722.5
712	2725	Matthew McConaugl	1718.5
315	2968	Brad Pitt	1641.5
733	2536	Henry Cavill	1634.5
406	2814	Christo Jivkov	1610
478	2685	Christian Bale	1581.5
401	2741	Tom Cruise	1571
775	2326	Michael Fassbender	1550.5
446	2618	Naomi Watts	1532
275	2789	Steve Bastoni	1532
446	2618	Naomi Watts	1532
446	2618	Naomi Watts	1532



Result

The analysis of the movie dataset provided valuable insights into the movie industry, including the factors that contribute to the success of a movie, the most popular genres and actors, and the changes in audience preferences over time. The findings of this analysis can be used by movie studios and producers to make informed decisions about the production and marketing of movies, which can ultimately lead to greater success and profitability.