



# IMDB Movie Analysis

My Recommendation +



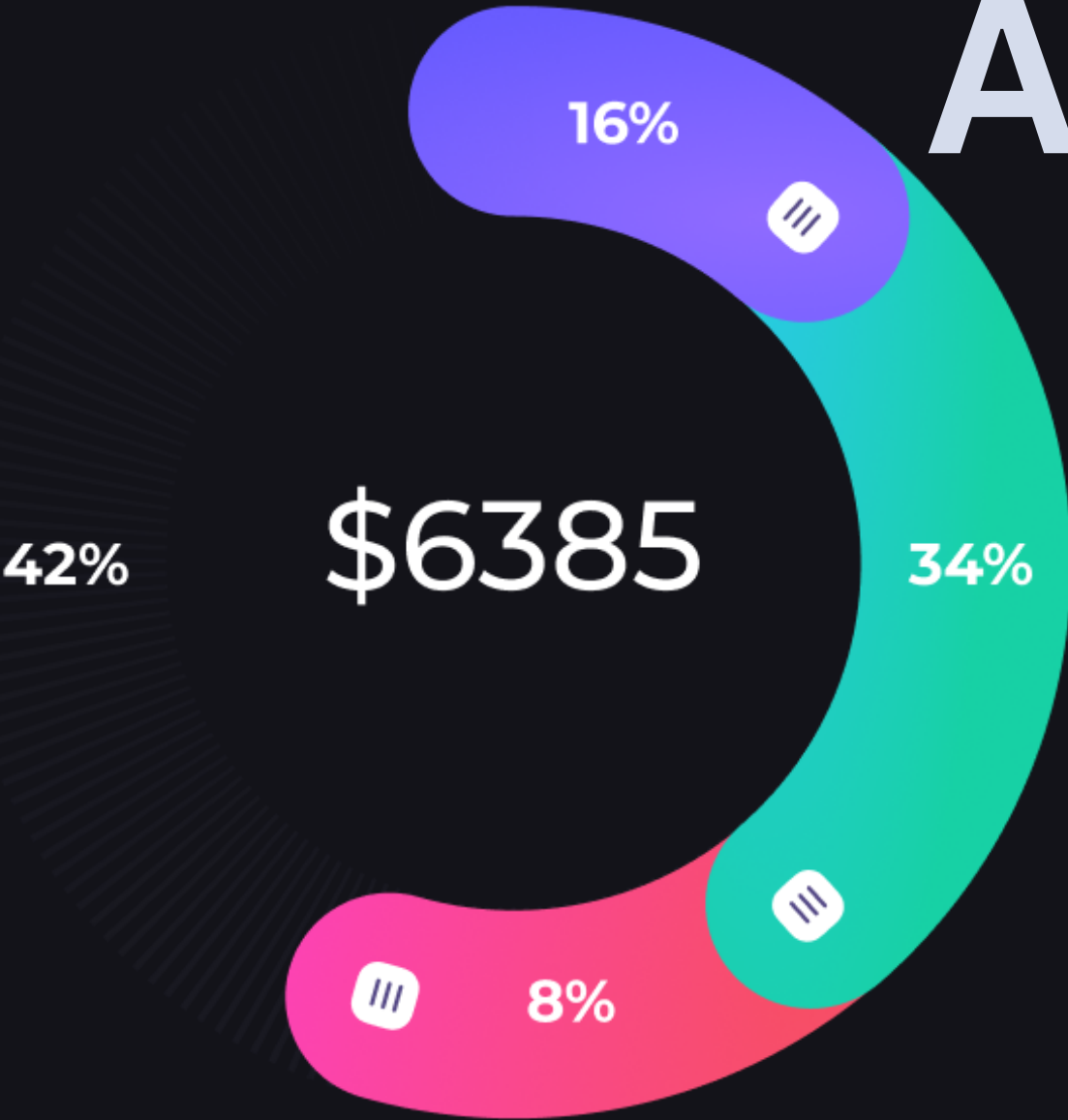
Parasite



Night At The Museum

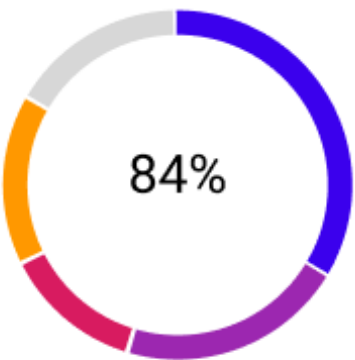


Deadpool



Movie Database

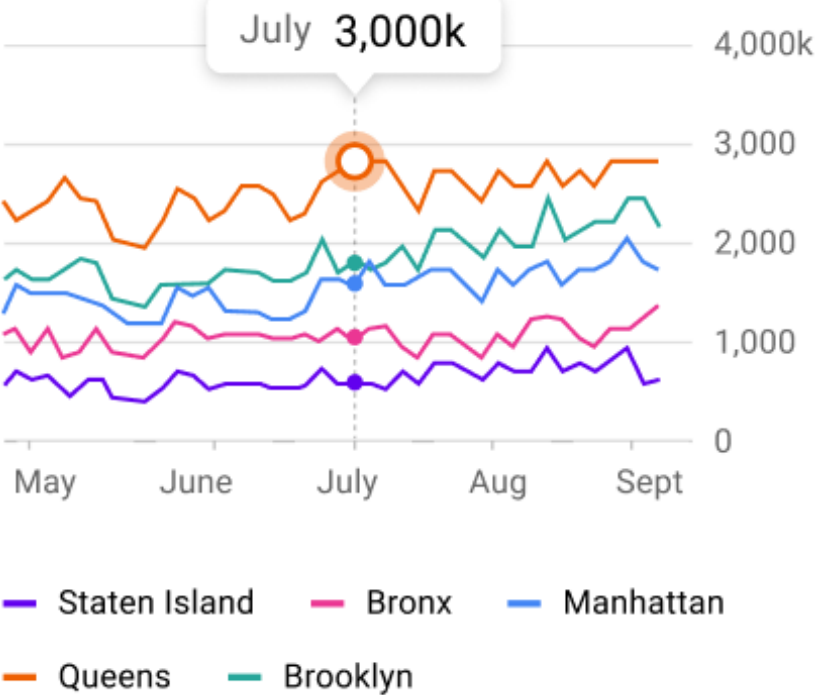
26.98



Photos & Videos	11.5
Music & Audio	7.15
Games	2.5
Movies & TV	5.83

Turnover

Estimated by borough



by Suraj Beloshe  
Powered by



# Project Description

1. The project involves analyzing a movie dataset using various data analysis techniques.
2. There are some tasks to perform like cleaning the data, finding movies with the highest profit, identifying the top 250 movies, finding the best directors, and identifying popular genres.
3. Additionally, the project involves creating new columns for some specific actors and combining them to identify the mean of the num\_critic\_for\_reviews and num\_users\_for\_review and identify the actors which have the highest mean.
4. The project also involves observing the change in number of voted users over decades using a bar chart
5. The main objective is to derive insights and trends from the dataset using data analysis techniques.

# Approach



1

Understand  
Dataset

2

Evaluate  
Insights

3

Use specific  
columns

4

Perform  
Analysis

5

Note Down  
insights

# Tech-Stack Used

## Microsoft 365 Excel

- **Familiarity:** MS Excel is a widely used software for data analysis and management. You might already be familiar with its interface and functions, which can make the project easier to complete.
- **Accessibility:** MS Excel 365 is a cloud-based software that can be accessed from anywhere with an internet connection. This means that you can work on your project from different devices and collaborate with others easily.
- **Data analysis tools:** MS Excel 365 provides a variety of data analysis tools, such as sorting, filtering, pivot tables, and charts, which can help you derive insights and trends from your data.
- **Integration with other Microsoft tools:** MS Excel 365 can be integrated with other Microsoft tools such as Power BI, SharePoint, and Teams, which can enhance your data analysis and presentation capabilities.



# Insights

## A. Cleaning the data:

- Data cleaning is an **essential step** in ensuring data consistency and accuracy. Removing any duplicate records in the dataset is crucial for this purpose.
- To ensure the integrity of complete data analysis, any missing values or null data should be replaced with appropriate values such as mean, median, or mode.
- The given dataset had **28 columns** and **5044 rows**. total **1,41,232 cells**
- There was a **special character** in the Movie Title column that was removed to maintain accuracy & readability.
- The dataset had **2697** blank cells, which could have had a significant impact on the findings. Therefore, it was crucial to remove this blank data.
- Using the "Go to Special" function, we identified the blank rows and replaced them with "TRUE".
- I used an **advanced filter with "TRUE"** to subtract the required data and moved it to another sheet, which we then converted into a table.
- Finally, I identified and removed duplicate values. In total, **33 duplicate** values were found, leaving **3723** unique rows.

FileHomeDeveloperInsertPage LayoutFormulasDataReviewViewAutomateHelp

UndoRedo

Clipboard

Font

Paragraph

Alignment

Number

Styles

Conditional Formatting

Format as Table

Cell Styles

Insert

Delete

Format

AutoSum

Filter

Sort & Filter

Find & Select

Analyze Data

Comments

Share

IMDB Movies analysis

Search

Suraj Beloshe

J15

fx

Action|Adventure|Sci-Fi

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
4	color	director_n	num_critics	duration	director_f	actor_3_f	actor_2_n	actor_1_f	gross	genres	actor_1_n	movie_title	num_voted	cast_total	actor_3_n	facnumbs	plot_keyw	movie_lm	num_user	language	country	content_r	budget	tit
5	Color	James Cam	723	178	0	855	Joel David	1000	7.61E+08	Action Ad CCH Pound	Avatar	886204	4834	Wes Studi	0	Avatar	1st http://www	3054	English	USA	PG-13	2.37E+08		
6	Color	Gore Verbi	302	169	563	1000	Orlando Bl	40000	3.09E+08	Action Ad Johnny De	Pirates of t	471220	48350	Jack Daver	0	goddess n	http://www	1238	English	USA	PG-13	3E+08		
7	Color	Sam Mend	602	148	0	161	Rory Kinne	11000	2E+08	Action Ad Christoph	Spectre	275868	11700	Stephanie	1	bomb esp	http://www	994	English	UK	PG-13	2.45E+08		
8	Color	Christophe	813	164	22000	23000	Christian B	27000	4.48E+08	Action Th Tom Hardy	The Dark K	1144337	106759	Joseph Go	0	deception	http://www	2701	English	USA	PG-13	2.5E+08		
9	TRUE	Doug Walk	TRUE	TRUE	131	TRUE	Rob Walke	131	TRUE	Document	Doug Walk	Star Wars:	8	143	TRUE	0	TRUE	http://www	TRUE	TRUE	TRUE	TRUE	TRUE	
10	Color	Andrew St	462	132	475	530	Samantha	640	73058679	Action Ad Daryl Saba	John Carte	212204	1873	Polly Walk	1	alien ame	http://www	738	English	USA	PG-13	2.64E+08		
11	Color	Sam Raimi	392	156	0	4000	James Fra	24000	3.37E+08	Action Ad J.K. Simmc	Spider-Ma	383056	46055	Kirsten Dui	0	sandman	http://www	1902	English	USA	PG-13	2.58E+08		
12	Color	Nathan Gr	324	100	15	284	Donna Ma	799	2.01E+08	Adventure	Brad Garre	Tangled	294810	2036	M.C. Gaini	1	17th centu	http://www	387	English	USA	PG	2.6E+08	
13	Color	Joss Whed	635	141	0	19000	Robert Do	26000	4.59E+08	Action Ad Chris Hem	Avengers:	462669	92000	Scarlett Jo	4	artificial	in http://www	1117	English	USA	PG-13	2.5E+08		
14	Color	David Yates	375	153	282	10000	Daniel Rad	25000	3.02E+08	Adventure	Alan Rickn	Harry Pott	321795	58753	Rupert Gri	3	blood boc	http://www	973	English	UK	PG	2.5E+08	
15	Color	Zack Snyder	673	183	0	2000	Lauren Coh	15000	3.3E+08	Action Ad Henry Cavi	Batman v	371639	24450	Alan D. Pui	0	based on c	http://www	3018	English	USA	PG-13	2.5E+08		
16	Color	Bryan Sing	434	169	0	903	Marlon Bro	18000	2E+08	Action Ad Kevin Spac	Superman	240396	29991	Frank Lang	0	crystal epi	http://www	2367	English	USA	PG-13	2.09E+08		
17	Color	Marc Forst	403	106	395	393	Mathieu A	451	1.68E+08	Action Ad Giancarlo	Quantum c	330784	2023	Rory Kinne	1	action nen	http://www	1243	English	UK	PG-13	2E+08		
18	Color	Gore Verbi	313	151	563	1000	Orlando Bl	40000	4.23E+08	Action Ad Johnny De	Pirates of t	522040	48486	Jack Daver	2	box office	http://www	1832	English	USA	PG-13	2.35E+08		
19	Color	Gore Verbi	450	150	563	1000	Ruth Wilsc	40000	89289910	Action Ad Johnny De	The Lone f	181792	45757	Tom Wikis	1	horse outl	http://www	711	English	USA	PG-13	2.15E+08		
20	Color	Zack Snyder	733	143	0	748	Christophe	15000	2.91E+08	Action Ad Henry Cavi	Man of Ste	548573	20495	Harry Lenr	0	based on c	http://www	2536	English	USA	PG-13	2.25E+08		
21	Color	Andrew Ac	258	150	80	201	Pierfrance	22000	1.42E+08	Action Ad Peter Dink	The Chron	149922	22697	Dami	4	brother bri	http://www	438	English	USA	PG	2.25E+08		
22	Color	Joss Whed	703	173	0	19000	Robert Do	26000	6.23E+08	Action Ad Chris Hem	The Averg	995415	87697	Scarlett Jo	3	alien invas	http://www	1722	English	USA	PG-13	2.2E+08		
23	Color	Rob Marsh	448	136	252	1000	Sam Claflir	40000	2.41E+08	Action Ad Johnny De	Pirates of t	370704	54083	Stephen Gi	4	blackbear	http://www	484	English	USA	PG-13	2.5E+08		
24	Color	Barry Sonr	451	106	188	718	Michael St	10000	1.79E+08	Action Ad Will Smith	Men in Bla	268154	12572	Nicole Sch	1	alien rim	http://www	341	English	USA	PG-13	2.25E+08		
25	Color	Peter Jack	422	164	0	773	Adam Brown	5000	2.55E+08	Adventure	Aidan Turn	The Hobbit	354228	9152	James Nes	0	army elf	http://www	802	English	New Zeala	PG-13	2.5E+08	
26	Color	Marc Webb	599	153	464	963	Andrew Gar	15000	2.62E+08	Action Ad Emma Stone	The Amazi	451803	28489	Chris Zylka	0	liard outl	http://www	1225	English	USA	PG-13	2.3E+08		
27	Color	Ridley Sco	343	156	0	738	William Hl	891	1.05E+08	Action Ad Mark Addy	Robin Hoo	211765	3244	Scott Grim	0	1190s arci	http://www	546	English	USA	PG-13	2E+08		
28	Color	Peter Jack	509	186	0	773	Adam Brown	5000	2.58E+08	Adventure	Aidan Turn	The Hobbit	483540	9152	James Nes	6	dwarf elf	http://www	951	English	USA	PG-13	2.25E+08	
29	Color	Chris Weit	251	113	129	1000	Eva Green	16000	70083519	Adventure	Christopher Lee	The Golden Co	149019	24106	Wristin Sco	2	children e	http://www	666	English	USA	PG-13	1.8E+08	
30	Color	Peter Jack	446	201	0	84	Thomas Kr	6000	2.18E+08	Action Ad Naomi Wa	King Kong	316018	7123	Evan Park	0	animal nar	http://www	2618	English	New Zeala	PG-13	2.07E+08		

Sheet1

A B C D E F G H I J K L M N O P Q R S T U V W

Ready

Sheet 1 of 7

Workbook Statistics

Num Lock

Accessibility: Investigate

30°C Haze

Search

TV

100%

ENG IN

2026

17-04-2023

AutoSave

IMDB8 Movies analysis

Search

Suraj Belose

FileHomeDeveloperInsertPage LayoutFormulasDataReviewViewAutomateHelpTable Design

ClipboardFontAlignmentNumberStyles

General

Conditional FormattingFormat as TableCell StylesInsertDeleteFormat

Wrap TextMerge & Center

Percent Styles

CommentsShare

AutoSumFillClearSort & FilterFind & SelectAnalyze Data

UndoRedo

B1

fx

director\_name

	B	C	D	E	F	G	H	I	J	K	L	M	N	
	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross	genres	actor_1_name	movie_title	num_voted_users	cast_total_fac	
1	James Cameron	723	178	0	855	Joel David Moore	1000	7.61E+08	Action Ad CCH Pounder	Avatar		886204		
2	Gore Verbinski	302	169	563	1000	Orlando Bloom	40000	3.09E+08	Action Ad Johnny Depp	Pirates of the C		471220		
3	Sam Mendes	602	148	0	161	Rory Kinnear	11000	2E+08	Action Ad Christoph Waltz	Spectre		275868		
4	Christopher Nolan	813	164	22000	23000	Christian Bale	27000	4.48E+08	Action Th Tom Hardy	The Dark Knigh		1144337		
5	Andrew Stanton	462	132	475	530	Samantha Morton	640	73058679	Action Ad Daryl Sabara	John Carter		212204		
6	Sam Raimi	392	156	0	4000	James Franco	24000	3.37E+08	Action Ad J.K. Simmons	Spider-Man 3		383056		
7	Nathan Greno	324	100	15	284	Donna Murphy	799	2.01E+08	Adventure	Brad Garrett	Tangled		294810	
8	Joss Whedon	635	141	0	19000	Robert Downey Jr	26000	4.59E+08	Action Ad Chris Hemsworth	The Avengers		462669		
9	David Yates	375	153	282	10000	Daniel Radcliffe	25000	3.02E+08	Adventure	Alan Rickman	Harry Potter ar		321795	
10	Zack Snyder	673	183	0	2000	Lauren Cohan	15000	3.3E+08	Action Ad Henry Cavill	Batman v Supe		371639		
11	Bryan Singer	434	169	0	903	Marlon Brando	18000	2E+08	Action Ad Kevin Spacey	Superman Ret.		240396		
12	Marc Forster	403	106	395	393	Mathieu Amalric	451	1.68E+08	Action Ad Giancarlo Giannini	Quantum of Sc		330784		
13	Gore Verbinski	313	151	563	1000	Orlando Bloom	40000	4.23E+08	Action Ad Johnny Depp	Pirates of the C		522040		
14	Gore Verbinski	450	150	563	1000	Ruth Wilson	40000	89289910	Action Ad Johnny Depp	The Lone Rang		181792		
15	Zack Snyder	733	143	0	748	Christopher Melor	15000	2.91E+08	Action Ad Henry Cavill	Man of Steel		548573		
16	Andrew Adamson	258	150	80	201	Pierfrancesco Fav	22000	1.42E+08	Action Ad Peter Dinklage	The Chronicles		149922		
17	Joss Whedon	703	173	0	19000	Robert Downey Jr	26000	6.23E+08	Action Ad Chris Hemsworth	The Avengers		995415		
18	Rob Marshall	448	136	252	1000	Sam Claflin	40000	2.41E+08	Action Ad Johnny Depp	Pirates of the C		370704		
19	Barry Sonnenfeld	451	106	188	718	Michael Stuhlbarg	10000	1.79E+08	Action Ad Will Smith	Men in Black 3		268154		
20	Peter Jackson	422	164	0	773	Adam Brown	5000	2.55E+08	Adventure	Aidan Turner	The Hobbit: Th		354228	
21	Marc Webb	599	153	464	963	Andrew Garfield	15000	2.62E+08	Action Ad Emma Stone	The Amazing S		451803		
22	Ridley Scott	343	156	0	738	William Hurt	891	1.05E+08	Action Ad Mark Addy	Robin Hood		211765		
23	Peter Jackson	509	186	0	773	Adam Brown	5000	2.58E+08	Adventure	Aidan Turner	The Hobbit: Th		483540	
24	Chris Weitz	251	113	129	1000	Eva Green	16000	70083519	Adventure	Christopher Lee	The Golden Co		149019	
25	Peter Jackson	446	201	0	84	Thomas Kretschm	6000	2.18E+08	Action Ad Naomi Watts	King Kong		316018		
26	James Cameron	315	194	0	794	Kate Winslet	29000	6.59E+08	Drama Ro Leonardo DiCapri	Titanic		793059		

Sheet1

A

B

C

D

E

F

G

H

I

J

K

L

M

N

Ready

30°C Haze

Search

100%

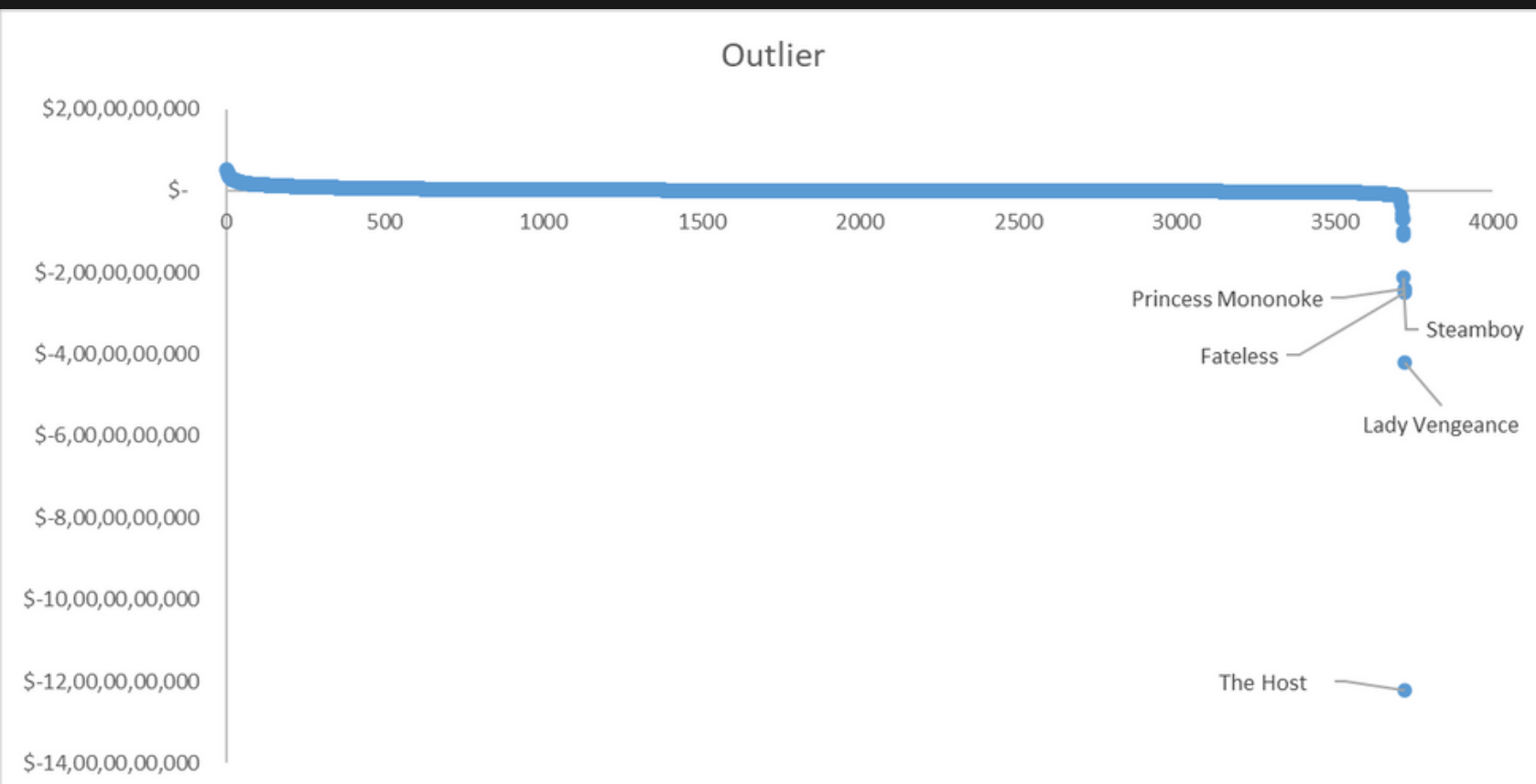
2026

17-04-2023

# Insights

## B. Movies with highest profit

movie_title	budget	gross	Profit
Avatar	\$ 23,70,00,000	\$ 76,05,05,847	\$ 52,35,05,847
Jurassic World	\$ 15,00,00,000	\$ 65,21,77,271	\$ 50,21,77,271
Titanic	\$ 20,00,00,000	\$ 65,86,72,302	\$ 45,86,72,302
Star Wars: Episode IV - A New Hope	\$ 1,10,00,000	\$ 46,09,35,665	\$ 44,99,35,665
E.T. the Extra-Terrestrial	\$ 1,05,00,000	\$ 43,49,49,459	\$ 42,44,49,459
The Avengers	\$ 22,00,00,000	\$ 62,32,79,547	\$ 40,32,79,547
The Lion King	\$ 4,50,00,000	\$ 42,27,83,777	\$ 37,77,83,777
Star Wars: Episode I - The Phantom Menace	\$ 11,50,00,000	\$ 47,45,44,677	\$ 35,95,44,677
The Dark Knight	\$ 18,50,00,000	\$ 53,33,16,061	\$ 34,83,16,061
The Hunger Games	\$ 7,80,00,000	\$ 40,79,99,255	\$ 32,99,99,255
Deadpool	\$ 5,80,00,000	\$ 36,30,24,263	\$ 30,50,24,263



- In order to determine the most profitable movies, I first needed to calculate the profits. Using a simple arithmetic formula, I calculated the profits for each movie.
- Next, I sorted the table with a "largest to smallest" filter and selected the **top 10 movies with the highest profits**.
- However, I discovered an outlier in our data. An outlier is an observation that significantly differs from other observations in a dataset.
- To identify the outliers, **we used a scatter chart** in Excel to plot the data points and look for points that are far away from the general trend of the data.
- Outliers can be easily visualized as points that are outside the main cluster of data points on the scatter chart.
- In our data, we found that the movie **'The Host'** had a profit of **\$-12,21,32,98,588.00**, which was significantly lower than the other movies and thus stood out as an outlier.
- There were four other movies that also stood out as outliers, as shown in the scatter chart.



# Insights

## C. Top 250

Rank	movie_title	num_voted _users	language	imdb_ score
1	The Shawshank Redemption	1689764	English	9.3
2	The Godfather	1155770	English	9.2
3	The Dark Knight	1676169	English	9
248	The Last of the Mohicans	113068	English	7.8
249	Apocalypto	236000	Maya	7.8
250	Fantastic Mr. Fox	139114	English	7.8

Rank	movie_title	num_voted _users	language	imdb_ score
8	The Good, the Bad and the Ugly	503509	Italian	8.9
19	City of God	533200	Portuguese	8.7
20	Seven Samurai	229012	Japanese	8.7
219	Letters from Iwo Jima	132149	Japanese	7.9
223	Amour	70382	French	7.9
249	Apocalypto	236000	Maya	7.8

- I were instructed to find the top 250 movies based on their IMDB score, with a minimum of 25,000 votes.
- Using an advanced filter, I separated the movies with 25,000 or more voted users from the rest.
- I then used a **LARGE & SEQUENCE formula** to get the top 250 movies based on their IMDB score.
- However, I faced the challenge of duplicate values.
- To solve this, I created a helper column to get unique values.
- Finally, I used an "XLOOKUP" function to obtain all the top 250 movies as required.
- I also wanted to find foreign language movies from the top 250 list.
- I used an **advanced filter with the criterion "not equal to English"** to extract all the foreign language movies from the top 250.

# Insights

## D. Best Directors

Sr. N	director_name	imdb_score
1	Akira Kurosawa	8.70
2	Charles Chaplin	8.60
3	Tony Kaye	8.60
4	Alfred Hitchcock	8.50
5	Damien Chazelle	8.50
6	Majid Majidi	8.50
7	Ron Fricke	8.50
8	Christopher Nolan	8.43
9	Sergio Leone	8.43
10	Asghar Farhadi	8.40
11	Richard Marquand	8.40
12	Billy Wilder	8.30
13	Fritz Lang	8.30
14	Lee Unkrich	8.30
15	Lenny Abrahamson	8.30
16	Hayao Miyazaki	8.23
17	Pete Docter	8.23
18	Elia Kazan	8.20
19	George Roy Hill	8.20
20	Joshua Oppenheimer	8.20

- To determine the best directors based on IMDb scores, I separated the two columns and grouped them as per the instructions.
- I used a **pivot table to group the data and find the mean** IMDb score for each director.
- By summarizing the values by average, we obtained a list of all directors with their respective mean IMDb scores.
- I then decided to find the top 20 directors from this list.
- To accomplish this, we used the **LARGE formula and the SEQUENCE formula**.
- With the help of a supporting column, I fetched the names of the directors and their IMDb scores.
- Further Sorted Data as instructed Alphabetically
- Finally, I obtained a list of the top 20 directors based on their mean IMDb scores.



# Insights

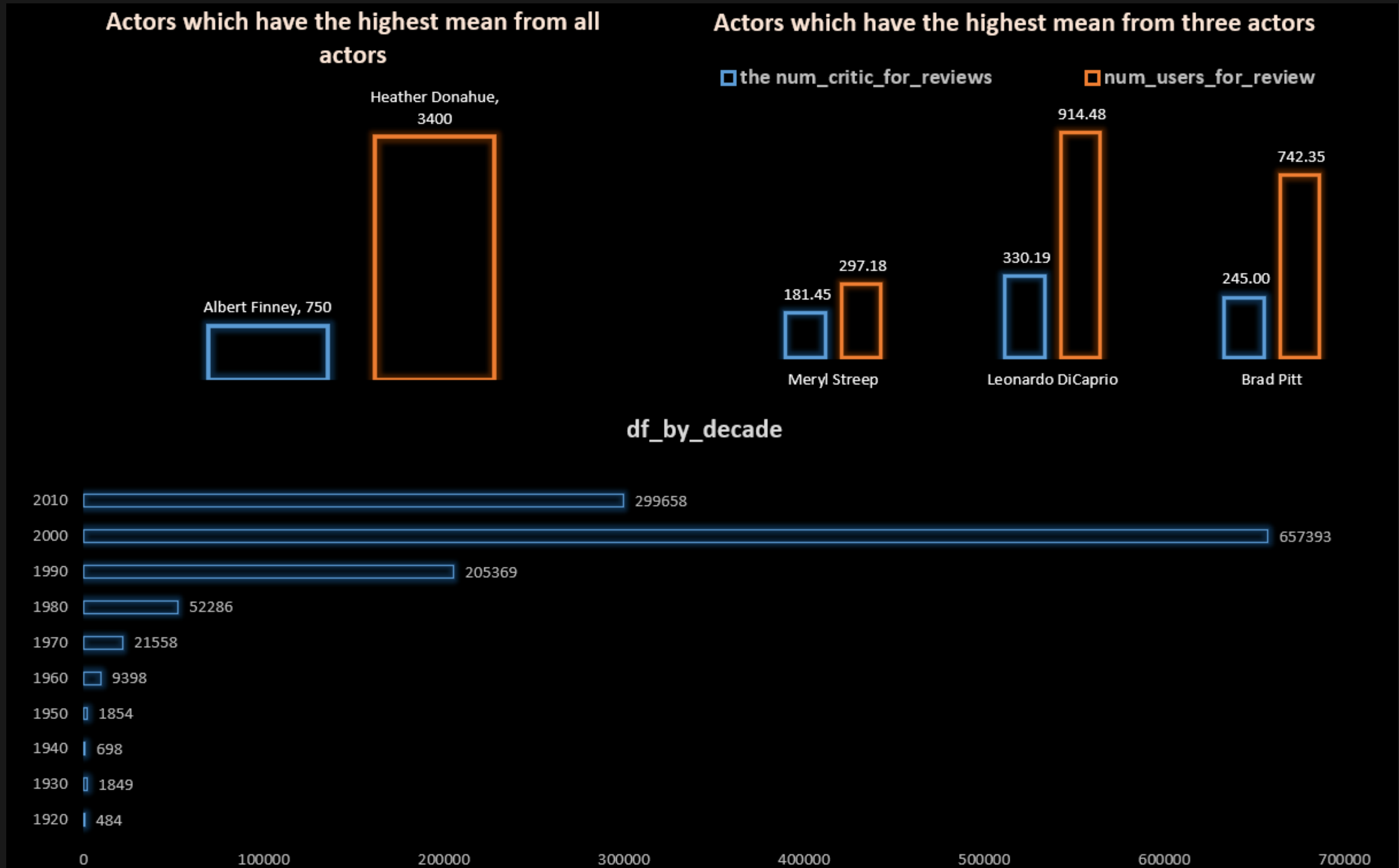
## E. Popular Genres

Popular Genres on the basis of Profit		
Sr. No.	genres	Sum of Profit
1	Comedy	3,00,43,60,840
2	Comedy Romance	2,69,28,38,991
3	Action Adventure Fantasy Sci-Fi	2,39,47,17,100
Popular Genres on the basis of number of voted users		
Sr. No.	genres	Sum of Num_Voted Users
1	Action Adventure Sci-Fi	1,46,23,810
2	Drama	1,23,19,080
3	Comedy	1,11,12,881
Popular Genres on the basis of Average of imdb score		
Sr. No.	genres	Average of imdb_score
1	Adventure Animation Drama Family Musical	8.50
2	Crime Drama Fantasy Mystery	8.50
3	Adventure Animation Fantasy	8.40

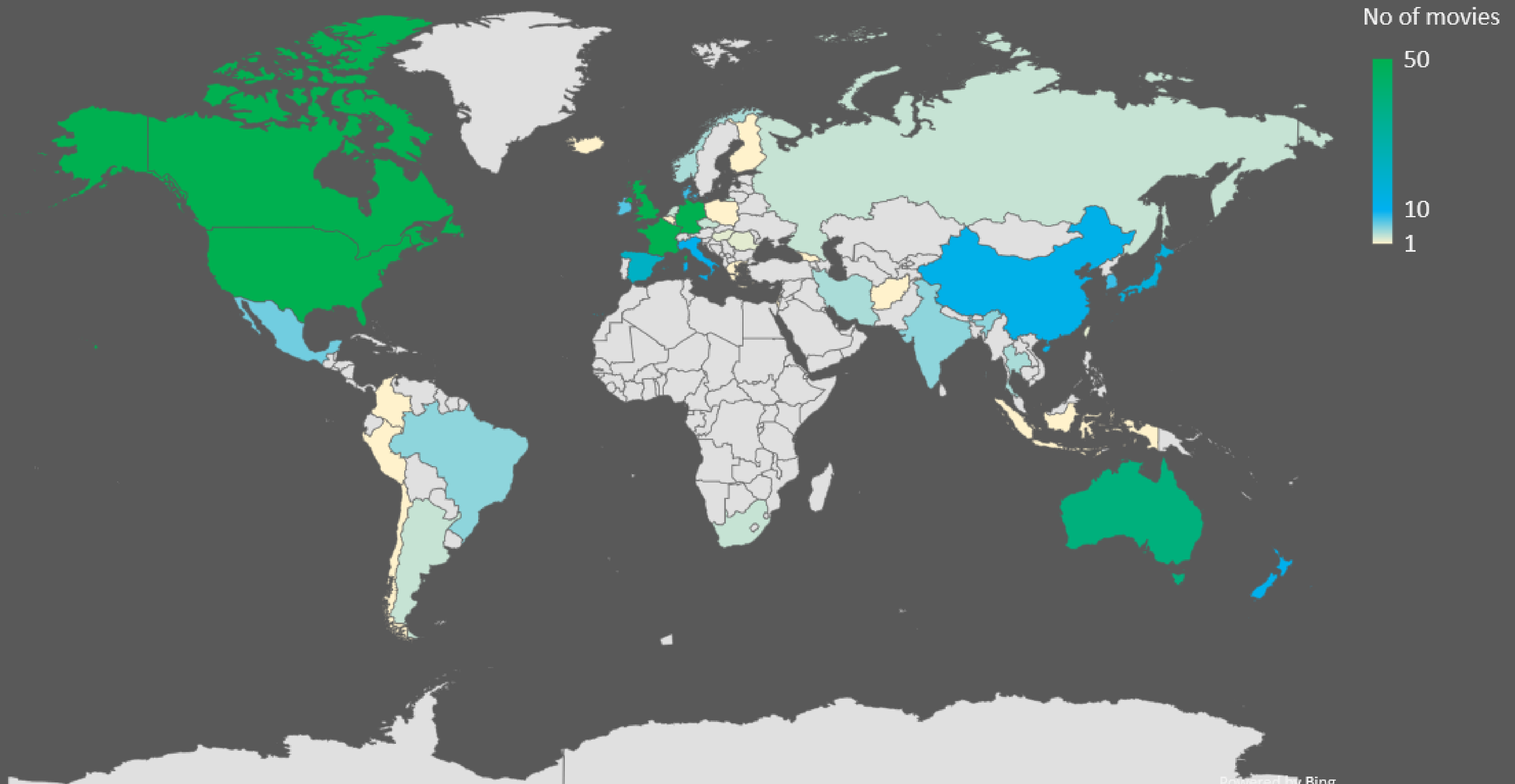
- For the first criterion, I calculated **movie profits** and used a pivot table with large sequence and XLOOKUP formulas to find the top three most popular genres based on earnings.
- For the second criterion, I used a pivot table to group data and find the **most voted genres**.
- For the third criterion, I used a unique value column and the sequence large and XLOOKUP formulas to calculate the **average IMDb** score for each genre.
- Using this method, I identified the top three most popular genres based on IMDb ratings.
- By using different criteria and data-driven methods, I were able to obtain a comprehensive list of the most popular genres in the movie industry.
- This information can be useful for movie producers and distributors in making decisions about which genres to invest in.

# Insights

## F. Charts



# Country wise listing of movies





# Result

1. Data cleaning is a crucial step in the data analysis process as it helps to ensure the accuracy and consistency of the data.
2. Data manipulation techniques like creating new columns, grouping the data by certain variables, and sorting the data can help in identifying patterns and insights.
3. Visualization techniques like scatter plots, bar charts and map charts can help in identifying trends, outliers and geographical representation in the data.
4. Grouping the data by certain variables can help in identifying trends and patterns in the data.
5. Using MS 365 to completed this project helps me to understand more about it and realize when some complex operation are taking more time than other language. Still it helps for data analysis with great accuracy and reliability.

Attachment:- Link for project file

**Thank You for  
your time....**