

PROJECT DESCRIPTION:

To analyse and give suitable solutions for the given queries by using Microsoft Excel.

APPROACH:

Understanding the given question and thinking logically and analytically to solve it.

TECH-STACK USED:

The software used in these projects is Microsoft Excel required solutions and to represent the data in graphical and chart format.

INSIGHTS:

By using Microsoft Excel for performing the given task it is necessary to analyse the given table and provide the suitable solution.

RESULT:

By solving the given problems, I have learned to think differently while shorting and aggregating datas.

1) Cleaning the data:: P This is one of the most important step 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

Use ctrl+g select the special button -> blank->ok

To select all the black cells

The screenshot shows a Microsoft Excel spreadsheet with a data table. The 'Go To Special' dialog box is open, and the 'Blanks' option is selected under the 'Select' tab. The data table has columns for movie details and actor information. The 'Go To Special' dialog box is used to select all blank cells in the selected range.

director_num	critic	duration	director_factor_3	actor_2_n	actor_1_fg	gross
James Car	723	178	0	855 Joel David	1000	7.61E+0
Gore Ver	302	169	563	1000 Orlando B	40000	3.09E+0
Sam Men	602	148	0	161 Rory Kinni	11000	2E+0
Christoph	813	164	22000	23000 Christian I	27000	4.48E+0
Doug Walk			131	Rob Walki	131	
Andrew Si	462	132	475	530 Samantha	640	7305867
Sam Raim	392	156	0	4000 James Fra	24000	3.37E+0
Nathan Gr	324	100	15	284 Donna Mu	799	2.01E+0
Joss Whec	635	141	0	19000 Robert Do	26000	4.59E+0
David Yate	375	153	282	10000 Daniel Rak	25000	3.02E+0
Zack Snyder	673	183	0	2000 Lauren Co	15000	3.3E+0
Bryan Sing	434	169	0	903 Marlon Br	18000	2E+0
Marc Forst	403	106	395	393 Mathieu A	451	1.68E+0
Gore Ver	313	151	563	1000 Orlando B	40000	4.23E+08
Gore Ver	450	150	563	1000 Ruth Wils	40000	89289910
Zack Snyder	733	143	0	748 Christoph	15000	2.91E+08
Andrew A	258	150	80	201 Pierfrance	22000	1.42E+08
Joss Whec	703	173	0	19000 Robert Do	26000	6.23E+08
Rob Mars	448	136	252	1000 Sam Clafl	40000	2.41E+08
Barry Son	451	106	188	718 Michael Si	10000	1.79E+08
Peter Jack	422	164	0	773 Adam Bro	5000	2.55E+08

Delete entire row by using ctrl+ - and by clicking the entire row radio button

IMDb_Movies - Excel (Product Activation Failed)

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW

Calibri 11 A⁺ Wrap Text General Conditional Formatting Format as Table Cell Styles Insert Delete Format AutoSum Fill Sort & Find & Filter Select Clear

Clipboard Font Alignment Number Styles Cells Editing

	A	B	C	D	E	F	G	H	I	
		director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross	genres
1	Color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross	genres
2	Color	James Cameron	723	178	0	855	Joel David Moore	1000	760505847	Action Adventure F
3	Color	Gore Verbinski	302	169	563	1000	Orlando Bloom	40000	309404152	Action Adventure F
4	Color	Sam Mendes	602	148	161	161	Rory Kinnear	11000	200074175	Action Adventure T
5	Color	Christopher Nolan	813	164	23000	23000	Christian Bale	27000	448130642	Action Thriller
6		Doug Walker					Rob Walker	131		Documentary
7	Color	Andrew Stanton	462	132	530	530	Samantha Morton	640	73058679	Action Adventure Si
8	Color	Sam Raimi	392	156	4000	4000	James Franco	24000	336330303	Action Adventure R
9	Color	Nathan Greno	324	100	284	284	Donna Murphy	799	200807262	Adventure Animatio
10	Color	Joss Whedon	635	141	19000	19000	Robert Downey Jr.	26000	458991599	Action Adventure Si
11	Color	David Yates	375	153	10000	10000	Daniel Radcliffe	25000	301956980	Adventure Family F
12	Color	Zack Snyder	673	183	0	2000	Lauren Cohan	15000	330249062	Action Adventure Si
13	Color	Bryan Singer	434	169	903	903	Marlon Brando	18000	200069408	Action Adventure Si
14	Color	Marc Forster	403	106	395	393	Mathieu Amalric	451	168368427	Action Adventure
15	Color	Gore Verbinski	313	151	563	1000	Orlando Bloom	40000	423032628	Action Adventure F
16	Color	Gore Verbinski	450	150	563	1000	Ruth Wilson	40000	89289910	Action Adventure W
17	Color	Zack Snyder	733	143	0	748	Christopher Meloni	15000	291021565	Action Adventure F
18	Color	Andrew Adamson	258	150	80	201	Pierfrancesco Favino	22000	141614023	Action Adventure F
19	Color	Joss Whedon	703	173	0	19000	Robert Downey Jr.	26000	623279547	Action Adventure Si
20	Color	Rob Marshall	448	136	252	1000	Sam Claflin	40000	241063875	Action Adventure F
21	Color	Barry Sonnenfeld	451	106	188	718	Michael Stuhlbarg	10000	179020854	Action Adventure C
22	Color	Peter Jackson	422	164	0	773	Adam Brown	5000	255108370	Adventure Fantasy
23	Color	Marc Webb	599	153	464	963	Andrew Garfield	15000	262030663	Action Adventure F

IMDb_Movies

2) 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?

Extrate the data from the given database

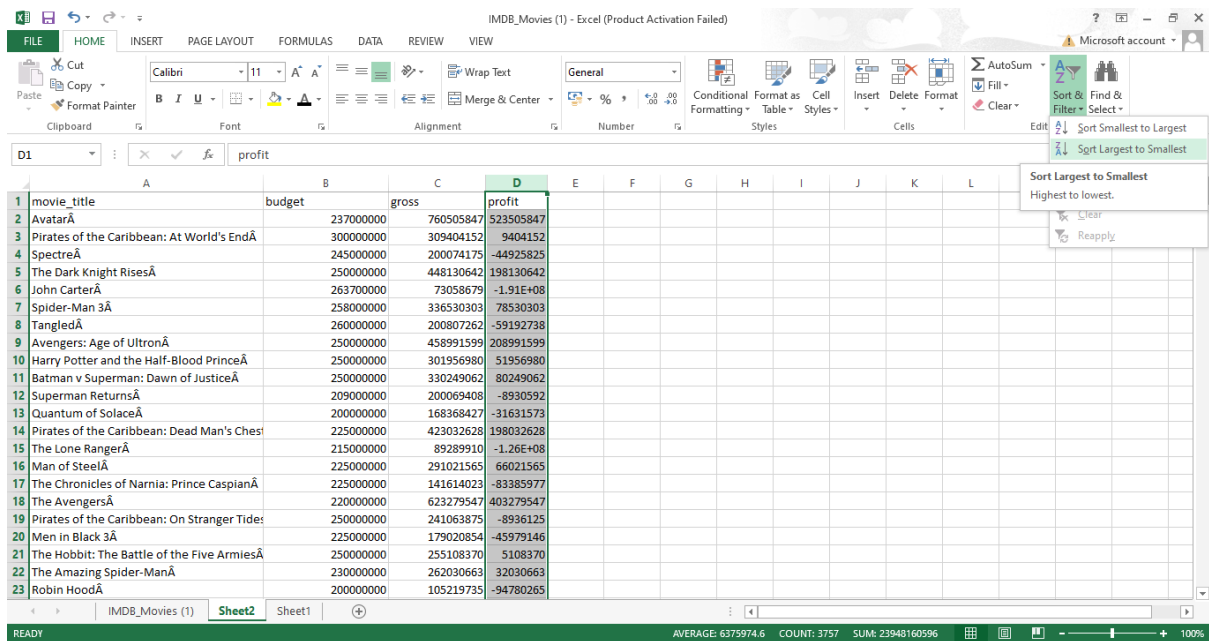
$$\text{Profit} = \text{gross} - \text{budget}$$

$$=c^2-b^2$$

Use the autofill option to apply that to the entire profit column

[illegible]

Sort the data from largest to smallest to get the movie with the highest profit



IMDB_Movies (1) - Excel (Product Activation Failed)

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW

Clipboard Font Alignment Number Styles Cells Editing

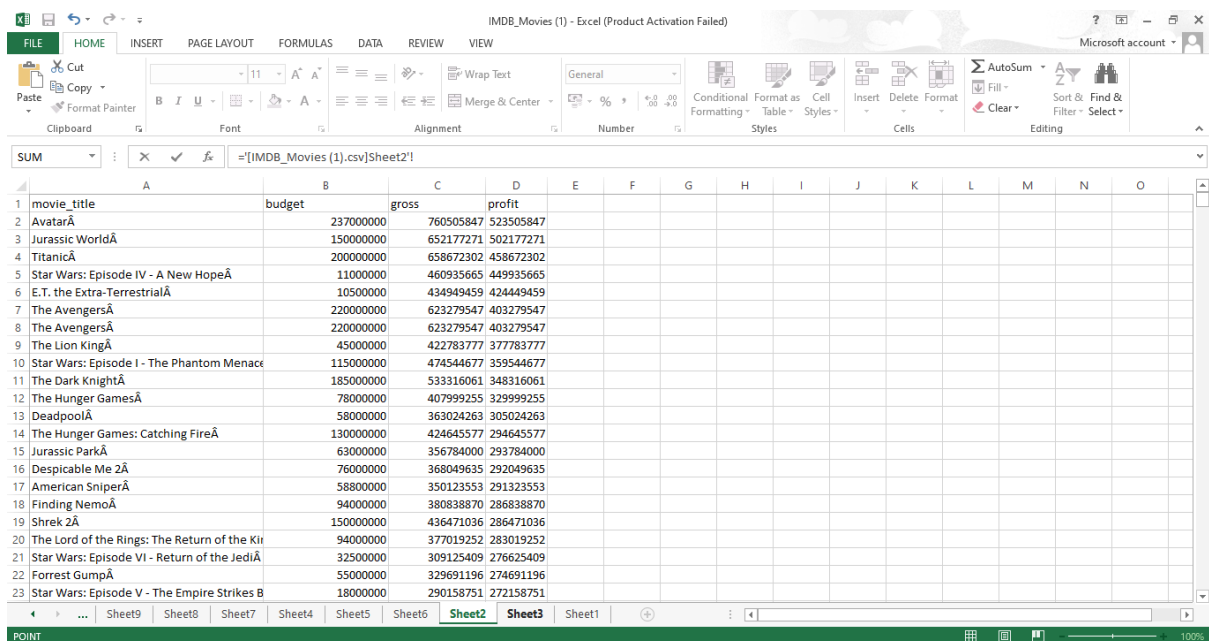
Sort & Filter Select

Sort Largest to Smallest
Highest to lowest.
Clear
Reapply

	A	B	C	D	E	F	G	H	I	J	K	L
	movie_title	budget	gross	profit								
1	Avatar	237000000	760505847	523505847								
2	Pirates of the Caribbean: At World's End	300000000	309404152	9404152								
3	Spectre	245000000	200074175	-44925825								
4	The Dark Knight Rises	250000000	448130642	198130642								
5	John Carter	263700000	73058679	-1.91E+08								
6	Spider-Man 3	258000000	336530303	78530303								
7	Tangled	260000000	200807262	-59192738								
8	Avengers: Age of Ultron	250000000	458991599	208991599								
9	Harry Potter and the Half-Blood Prince	250000000	301956980	51956980								
10	Batman v Superman: Dawn of Justice	250000000	330249062	80249062								
11	Superman Returns	209000000	200069408	-8930592								
12	Quantum of Solace	200000000	168368427	-31631573								
13	Pirates of the Caribbean: Dead Man's Chest	225000000	423032628	198032628								
14	The Lone Ranger	215000000	89289910	-1.26E+08								
15	Man of Steel	225000000	291021565	66021565								
16	The Chronicles of Narnia: Prince Caspian	225000000	141614023	-83385977								
17	The Avengers	220000000	623279547	403279547								
18	Pirates of the Caribbean: On Stranger Tides	250000000	241063875	-8936125								
19	Men in Black 3	225000000	179020854	-45979146								
20	The Hobbit: The Battle of the Five Armies	250000000	255108370	5108370								
21	The Amazing Spider-Man	230000000	262030663	32030663								
22	Robin Hood	200000000	105219735	-94780265								

IMDB_Movies (1) Sheet2 Sheet1

READY AVERAGE: 6375974.6 COUNT: 3757 SUM: 23948160596 100%



IMDB_Movies (1) - Excel (Product Activation Failed)

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW

Clipboard Font Alignment Number Styles Cells Editing

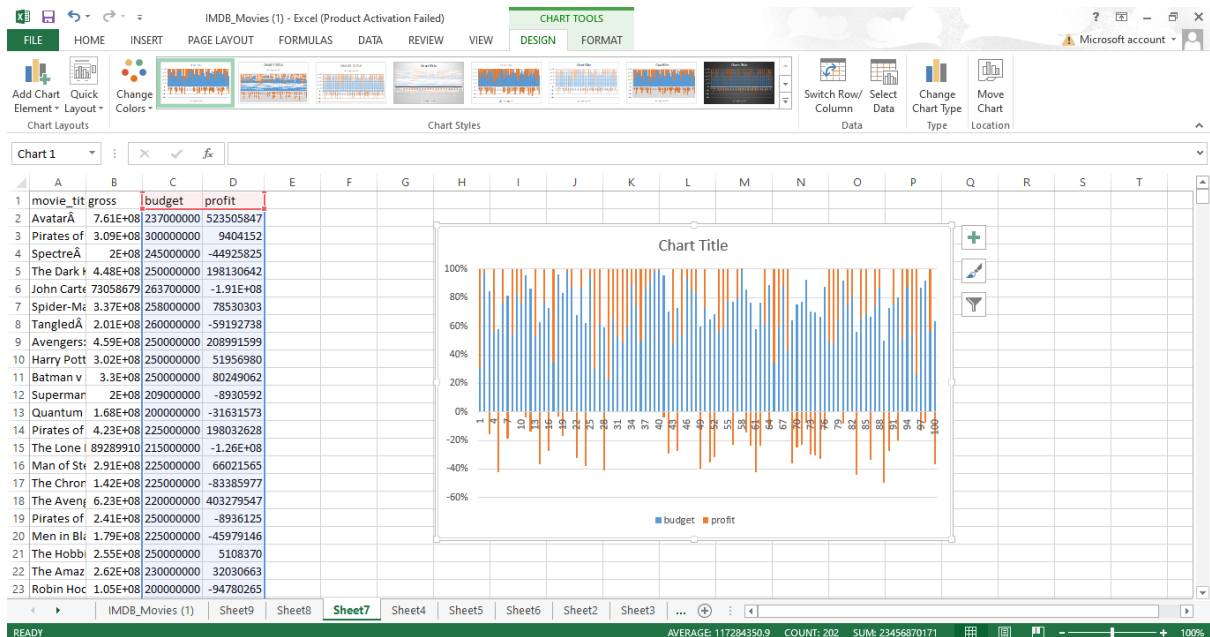
SUM =IMDB_Movies (1).csv!Sheet2!

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	movie_title	budget	gross	profit											
1	Avatar	237000000	760505847	523505847											
2	Jurassic World	150000000	652177271	502177271											
3	Titanic	200000000	658672302	458672302											
4	Star Wars: Episode IV - A New Hope	11000000	460935665	449935665											
5	E.T. the Extra-Terrestrial	10500000	434949459	424449459											
6	The Avengers	220000000	623279547	403279547											
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	180000000	290158751	272158751											

Sheet9 Sheet8 Sheet7 Sheet4 Sheet5 Sheet6 Sheet2 Sheet3 Sheet1

POINT 100%

Using the first 100 sample data with out sorting to analyse the budget and the profit margin



3) **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

Sorting imdb score from largest to the smallest so that it becomes easy in performing the calculation

IMDB_Movies - Excel (Product Activation Failed)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		imdb_score											
2		7.9	Avatar	886204									
3		7.1	Pirates of the Caribbean: At World's End	471220									
4		6.8	Spectre	275868									
5		8.5	The Dark Knight Rises	1144337									
6		6.6	John Carter	212204									
7		6.2	Spider-Man 3	383056									
8		7.8	Tangled	294810									
9		7.5	Avengers: Age of Ultron	462669									
10		7.5	Harry Potter and the Half-Blood Prince	321795									
11		6.9	Batman v Superman: Dawn of Justice	371639									
12		6.1	Superman Returns	240396									
13		6.7	Quantum of Solace	330784									
14		7.3	Pirates of the Caribbean: Dead Man's Chest	522040									
15		6.5	The Lone Ranger	181792									
16		7.2	Man of Steel	548573									
17		6.6	The Chronicles of Narnia: Prince Caspian	149922									
18		8.1	The Avengers	995415									
19		6.7	Pirates of the Caribbean: On Stranger Tides	370704									
20		6.8	Men in Black 3	268154									
21		7.5	The Hobbit: The Battle of the Five Armies	354228									
22		7	The Amazing Spider-Man	451803									
23		6.7	Robin Hood	211765									

Sort Largest to Smallest
Highest to lowest.

Clear
Reapply

IMDB_Movies Sheet4 Sheet3 Sheet2

AVERAGE: 6.453027523 COUNT: 3816 SUM: 24618.3

Creating a new column value and accepting the rows that have num_voted_users >25000

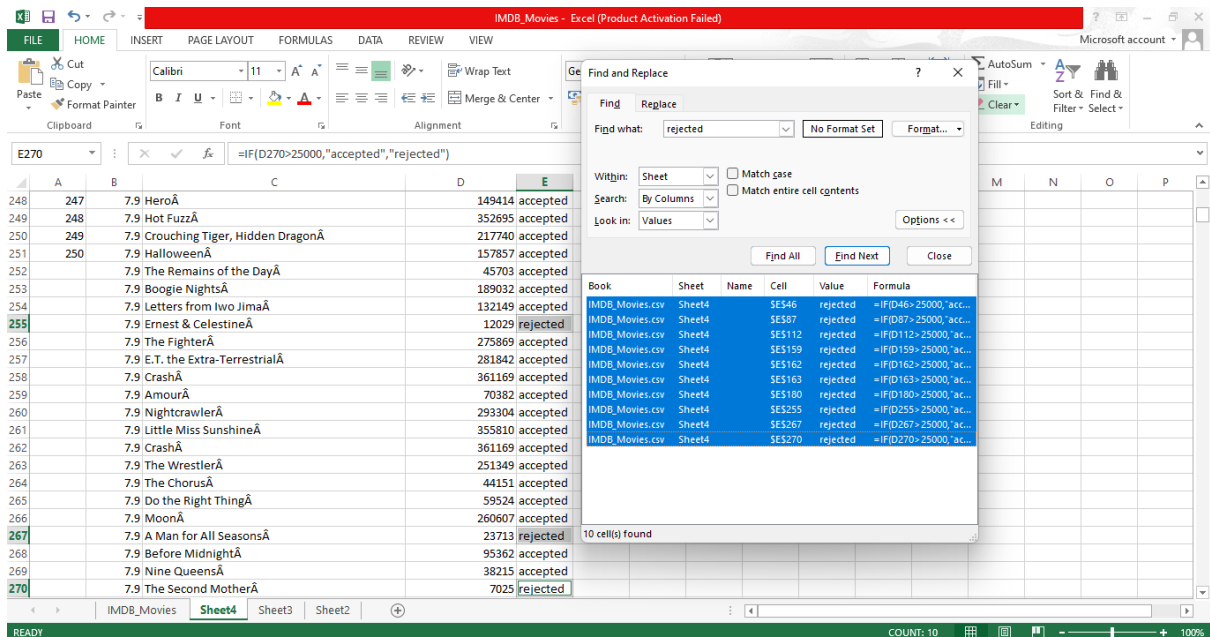
IMDB_Movies - Excel (Product Activation Failed)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	rank	imdb_score	movie_title	num_voted_users	value											
2	1	9.3	The Shawshank Redemption	1689764	=IF(D2>25000,"accepted","rejected")											
3	2	9.2	The Godfather	1155770	=IF(D3>25000,"accepted","rejected")											
4	3	9	The Dark Knight	1676169	accepted											
5	4	9	The Godfather: Part II	790926	accepted											
6	5	8.9	The Lord of the Rings: The Return of the King	1215718	accepted											
7	6	8.9	Schindler's List	865020	accepted											
8	7	8.9	Pulp Fiction	1324680	accepted											
9	8	8.9	The Good, the Bad and the Ugly	503509	accepted											
10	9	8.8	Inception	1468200	accepted											
11	10	8.8	The Lord of the Rings: The Fellowship of the Ring	1238746	accepted											
12	11	8.8	Fight Club	1347461	accepted											
13	12	8.8	Forrest Gump	1251222	accepted											
14	13	8.8	Star Wars: Episode V - The Empire Strikes Back	837759	accepted											
15	14	8.7	The Lord of the Rings: The Two Towers	1100446	accepted											
16	15	8.7	The Matrix	1217752	accepted											
17	16	8.7	Goodfellas	728685	accepted											
18	17	8.7	Star Wars: Episode IV - A New Hope	911097	accepted											
19	18	8.7	One Flew Over the Cuckoo's Nest	680041	accepted											
20	19	8.7	City of God	533200	accepted											
21	20	8.7	Seven Samurai	229012	accepted											
22	21	8.6	Interstellar	928227	accepted											
23	22	8.6	Saving Private Ryan	881236	accepted											

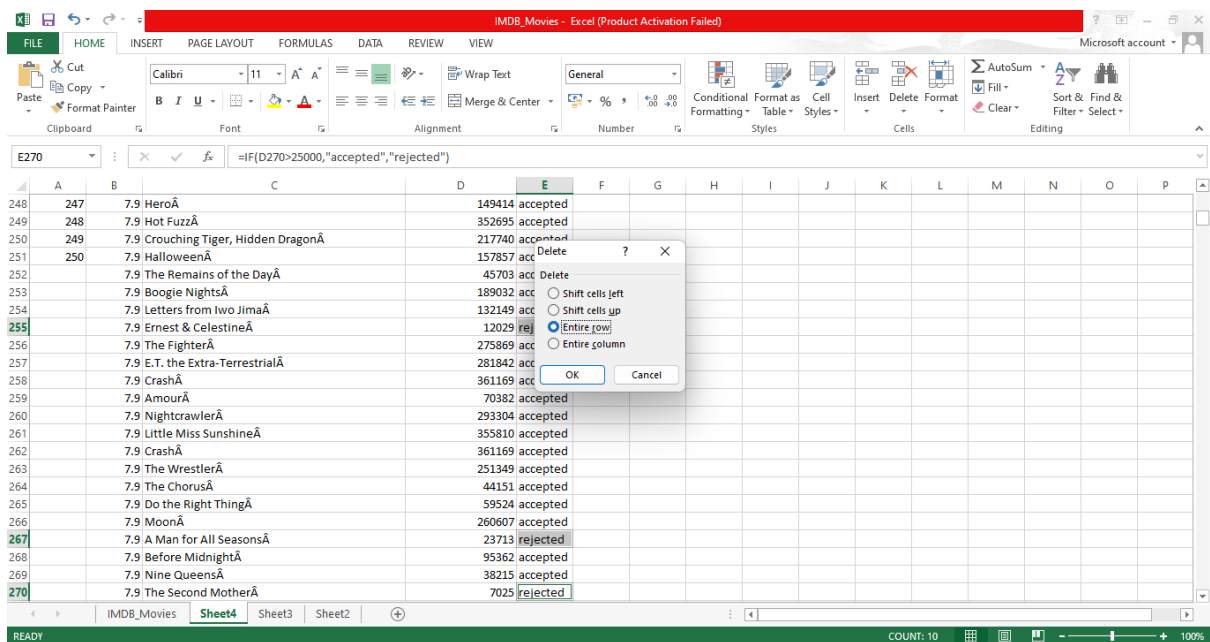
IMDB_Movies Sheet4 Sheet3 Sheet2

Selecting the cells with value "rejected"

By Using ctrl+f and find option

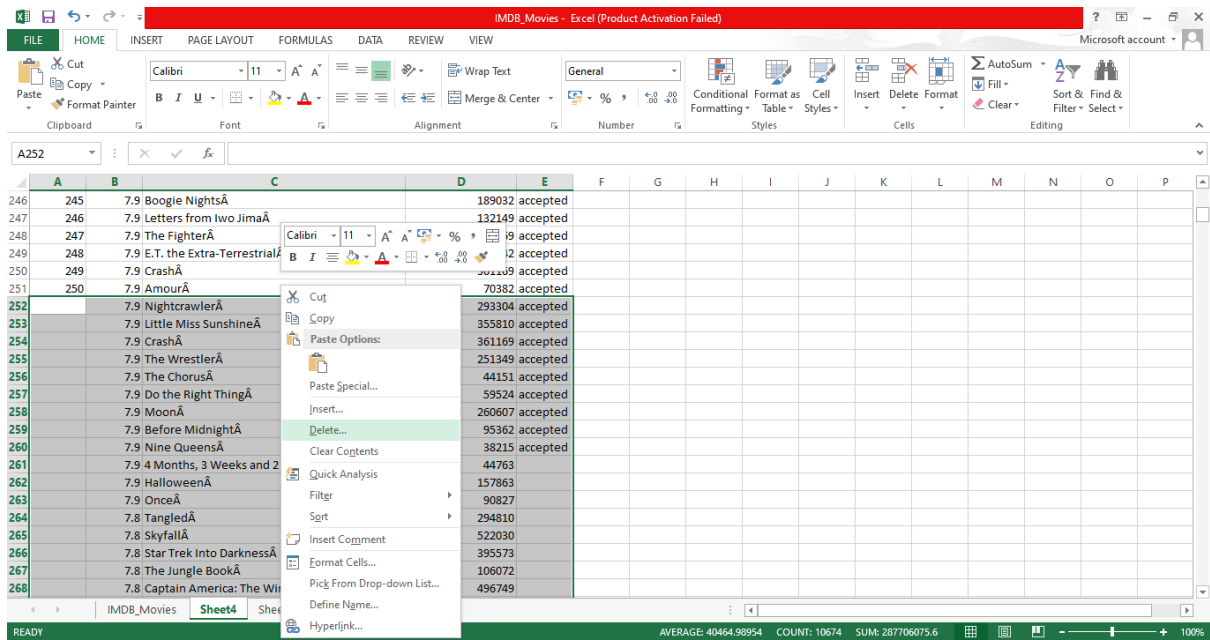


Deleting the selected rows by ctrl+ - and selecting entire row radio button



Ranking them from 1-250 by autofill option

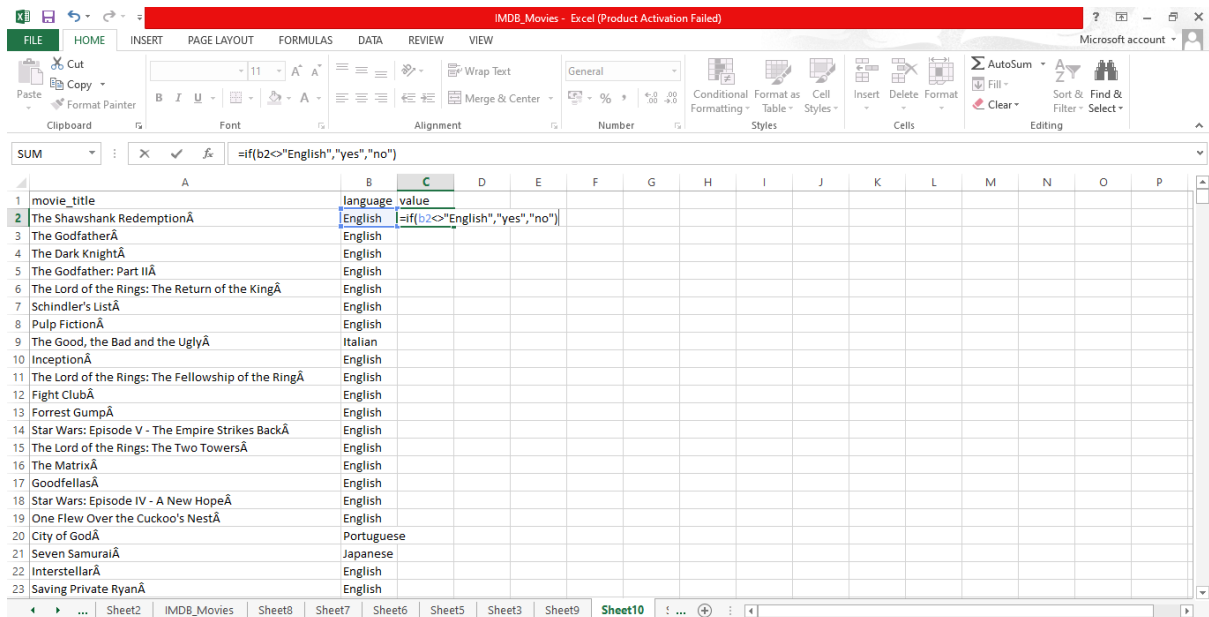
And deleting the excess data by selecting them with the range option



The Shawshank Redemption																			
rank	imdb_score	IMDb Top 250	num_voted_users																
2	1	9.3 The Shawshank Redemption	1689764																
3	2	9.2 The Godfather	1155770																
4	3	9 The Dark Knight	1676169																
5	4	9 The Godfather: Part II	790926																
6	5	8.9 The Lord of the Rings: The Return of the King	1215718																
7	6	8.9 Schindler's List	865020																
8	7	8.9 Pulp Fiction	1324680																
9	8	8.9 The Good, the Bad and the Ugly	503509																
10	9	8.8 Inception	1468200																
11	10	8.8 The Lord of the Rings: The Fellowship of the Ring	1238746																
12	11	8.8 Fight Club	1347461																
13	12	8.8 Forrest Gump	1251222																
14	13	8.8 Star Wars: Episode V - The Empire Strikes Back	837759																
15	14	8.7 The Lord of the Rings: The Two Towers	1100446																
16	15	8.7 The Matrix	1217752																
17	16	8.7 Goodfellas	728685																
18	17	8.7 Star Wars: Episode IV - A New Hope	911097																
19	18	8.7 One Flew Over the Cuckoo's Nest	680041																
20	19	8.7 City of God	533200																
21	20	8.7 Seven Samurai	229012																
22	21	8.6 Interstellar	928227																
23	22	8.6 Saving Private Ryan	881236																

For the top movies which are not English from the top 250 movies

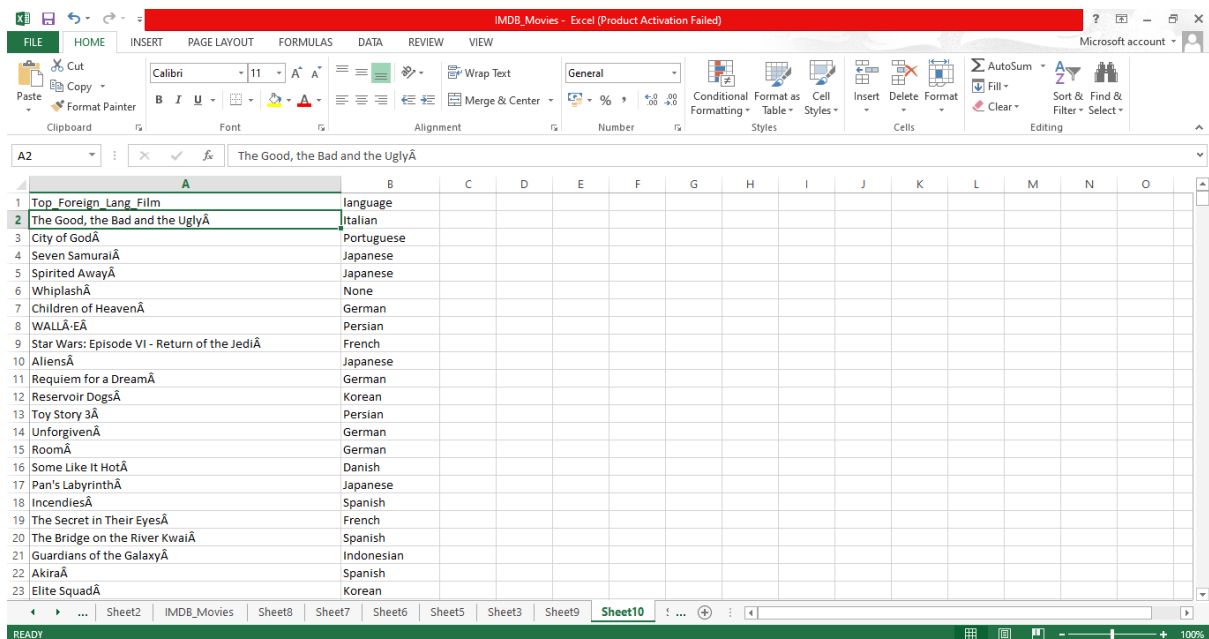
=if(<>"English,"yes","no")



The rows with English as its language are selected

By finding the rows that has no haveue using find&replace dialogbox

And deleting them using ctrl+ - command

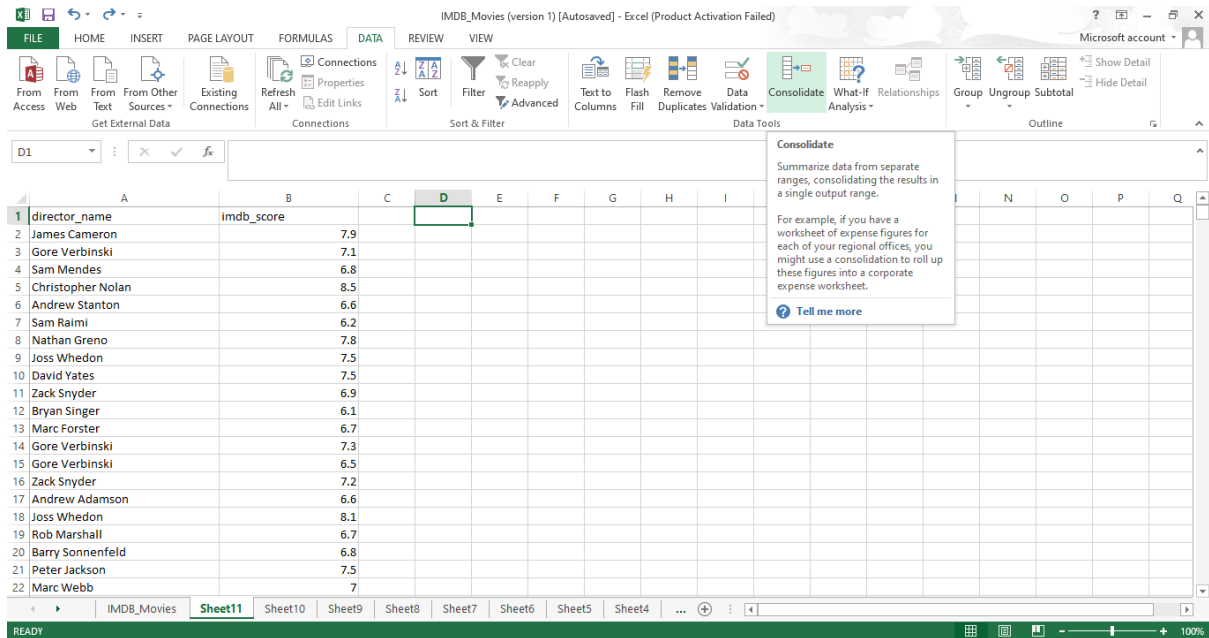


4) Best Directors: TGroup the column using the director_name column.

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

Using consolidate option find the average imbd_score of each director after groping them according to their names



IMDB_Movies (version 1) [Autosaved] - Excel (Product Activation Failed)

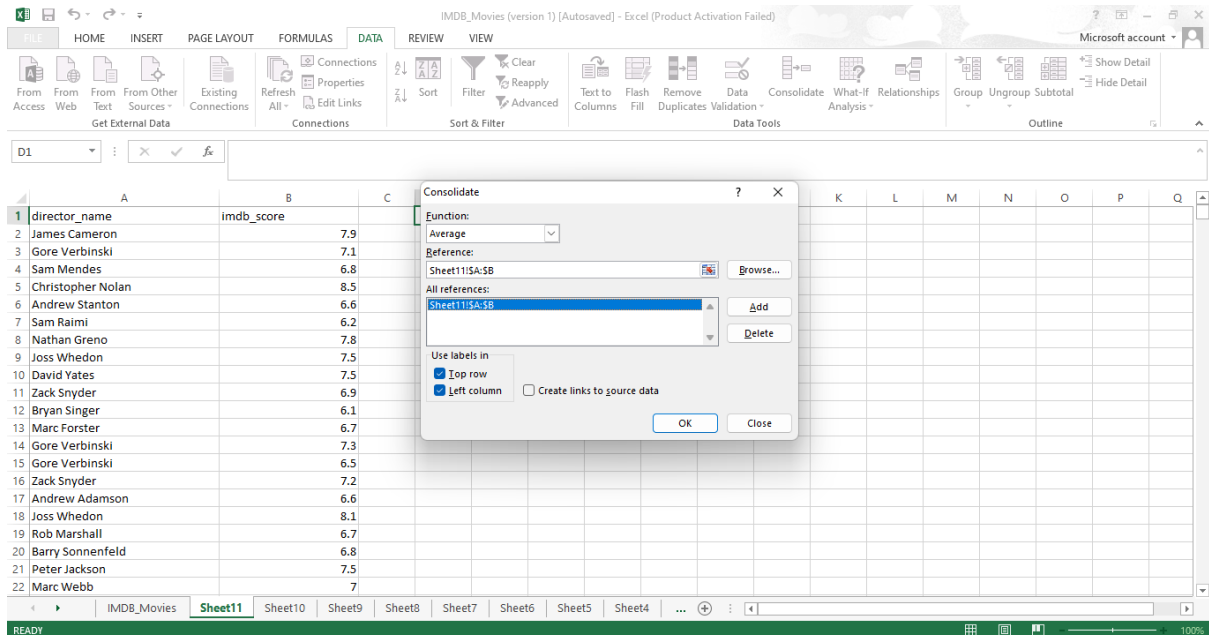
Consolidate

Summarize data from separate ranges, consolidating the results in a single output range.

For example, if you have a worksheet of expense figures for each of your regional offices, you might use a consolidation to roll up these figures into a corporate expense worksheet.

[Tell me more](#)

director_name	imdb_score
James Cameron	7.9
Gore Verbinski	7.1
Sam Mendes	6.8
Christopher Nolan	8.5
Andrew Stanton	6.6
Sam Raimi	6.2
Nathan Greno	7.8
Joss Whedon	7.5
David Yates	7.5
Zack Snyder	6.9
Bryan Singer	6.1
Marc Forster	6.7
Gore Verbinski	7.3
Gore Verbinski	6.5
Zack Snyder	7.2
Andrew Adamson	6.6
Joss Whedon	8.1
Rob Marshall	6.7
Barry Sonnenfeld	6.8
Peter Jackson	7.5
Marc Webb	7



IMDB_Movies (version 1) [Autosaved] - Excel (Product Activation Failed)

Consolidate

Function: Average

Reference: Sheet11!\$A:\$B

All references: Sheet11!\$A:\$B

Use labels in

☒ Top row

☒ Left column

☐ Create links to source data

OK Close

director_name	imdb_score
James Cameron	7.9
Gore Verbinski	7.1
Sam Mendes	6.8
Christopher Nolan	8.5
Andrew Stanton	6.6
Sam Raimi	6.2
Nathan Greno	7.8
Joss Whedon	7.5
David Yates	7.5
Zack Snyder	6.9
Bryan Singer	6.1
Marc Forster	6.7
Gore Verbinski	7.3
Gore Verbinski	6.5
Zack Snyder	7.2
Andrew Adamson	6.6
Joss Whedon	8.1
Rob Marshall	6.7
Barry Sonnenfeld	6.8
Peter Jackson	7.5
Marc Webb	7

IMDB_Movies (version 1) [Autosaved] - Excel (Product Activation Failed)

	A	B	C	D	E	F	G	H	I	J	K
1	director_name	imdb_score		director_name	imdb_score						
2	James Cameron	7.9		James Cameron	7.914285714						
3	Gore Verbinski	7.1		Gore Verbinski	6.985714286						
4	Sam Mendes	6.8		Sam Mendes	7.5						
5	Christopher Nolan	8.5		Christopher Nolan	8.425						
6	Andrew Stanton	6.6		Andrew Stanton	7.733333333						
7	Sam Raimi	6.2		Sam Raimi	6.85						
8	Nathan Greno	7.8		Nathan Greno	7.8						
9	Joss Whedon	7.5		Joss Whedon	7.925						
10	David Yates	7.5		David Yates	7.05						
11	Zack Snyder	6.9		Zack Snyder	7.175						
12	Bryan Singer	6.1		Bryan Singer	7.2875						
13	Marc Forster	6.7		Marc Forster	7.228571429						
14	Gore Verbinski	7.3		Andrew Adamson	7.15						
15	Gore Verbinski	6.5		Rob Marshall	6.6						
16	Zack Snyder	7.2		Barry Sonnenfeld	6.457142857						
17	Andrew Adamson	6.6		Peter Jackson	7.654545455						
18	Joss Whedon	8.1		Marc Webb	7.133333333						
19	Rob Marshall	6.7		Ridley Scott	7.070588235						
20	Barry Sonnenfeld	6.8		Chris Weitz	6.08						
21	Peter Jackson	7.5		Anthony Russo	7						
22	Marc Webb	7		Peter Berg	6.666666667						

READY AVERAGE: 6.310529393 COUNT: 1660 SUM: 10469.16826 100%

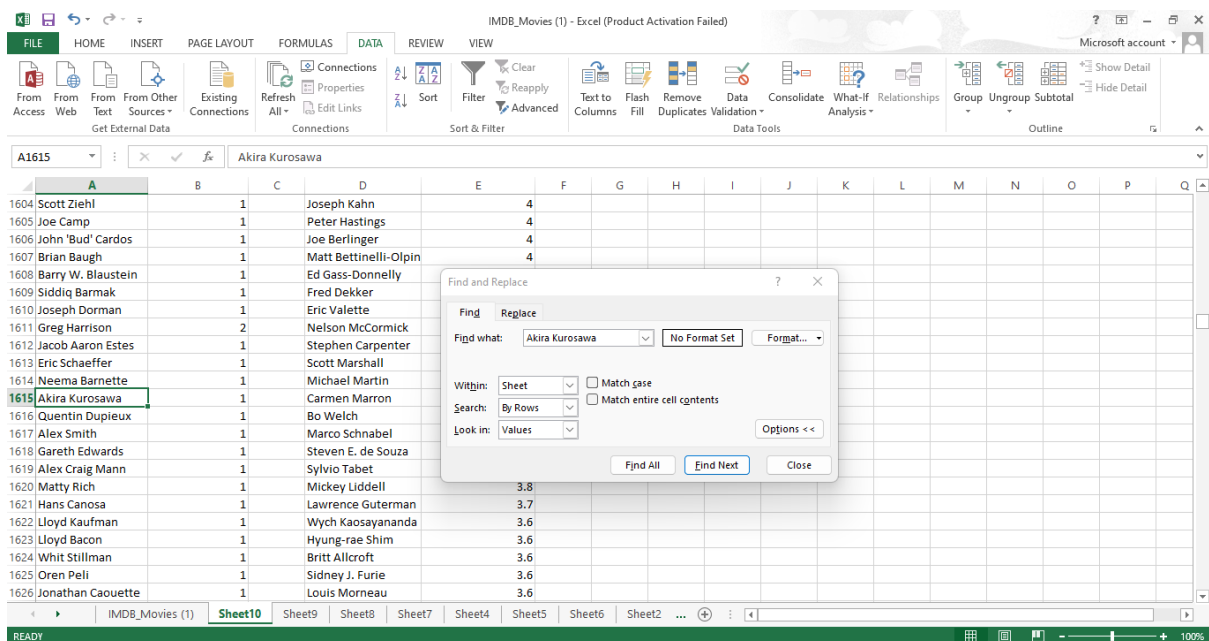
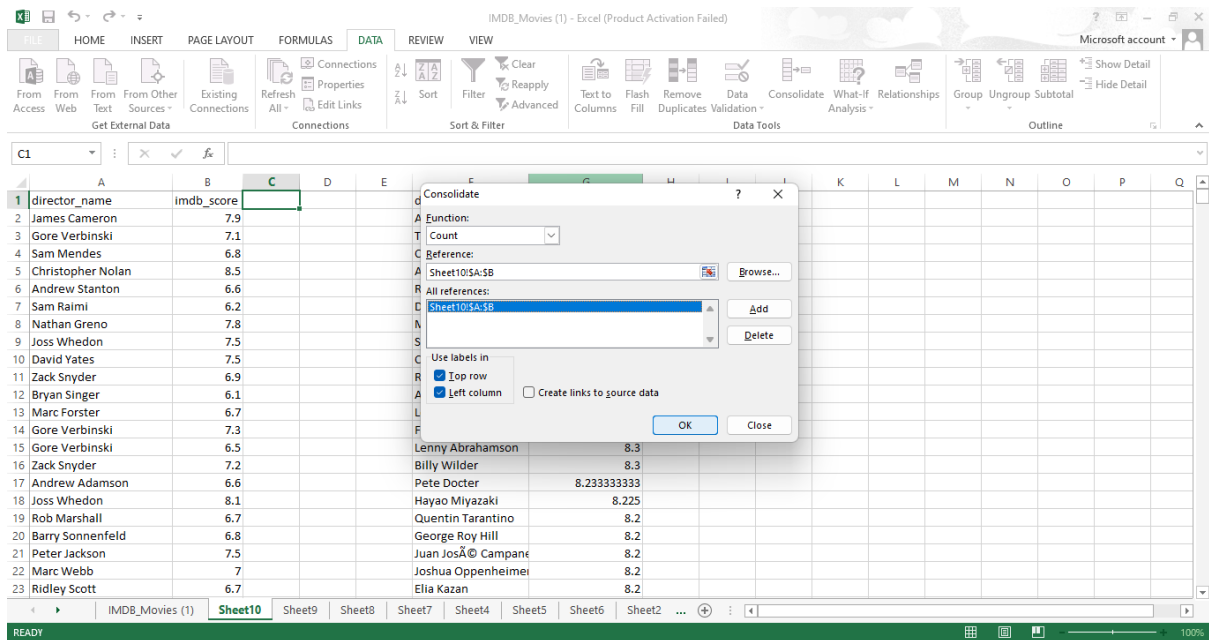
To find the top 10 sort it from the highest to lowest

IMDB_Movies (version 1) [Autosaved] - Excel (Product Activation Failed)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	rank	director_name	imdb_score															
2	1	Akira Kurosawa	8.7															
3	2	Charles Chaplin	8.6															
4	3	Tony Kaye	8.6															
5	4	Alfred Hitchcock	8.5															
6	5	Damien Chazelle	8.5															
7	6	Majid Majidi	8.5															
8	7	Ron Fricke	8.5															
9	8	Sergio Leone	8.433333333															
10	9	Christopher Nolan	8.425															
11	10	Asghar Farhadi	8.4															
12																		
13																		
14																		
15																		
16																		
17																		
18																		
19																		
20																		
21																		
22																		

READY 100%

use the count option grouping by director names to see whether any of them have done only one film



If u want to have only the directors who have done more than one film delete them

And sort them from a-> z if any two or more have the same value

IMDB_Movies (version 1) [Autosaved] - Excel (Product Activation Failed)

rank	director_name	imdb_score	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Tony Kaye	8.6													
2	Charles Chaplin	8.6													
3	Alfred Hitchcock	8.5													
4	Damien Chazelle	8.5													
5	Majid Majidi	8.5													
6	Ron Fricke	8.5													
7	Sergio Leone	8.433333333													
8	Christopher Nolan	8.425													
9	Richard Marquand	8.4													
10	Asghar Farhadi	8.4													

IMDB_Movies (version 1) [Autosaved] - Excel (Product Activation Failed)

rank	director_name	imdb_score	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Charles Chaplin	8.6															
2	Tony Kaye	8.6															
3	Alfred Hitchcock	8.5															
4	Damien Chazelle	8.5															
5	Majid Majidi	8.5															
6	Ron Fricke	8.5															
7	Sergio Leone	8.433333333															
8	Christopher Nolan	8.425															
9	Asghar Farhadi	8.4															
10	Richard Marquand	8.4															

5)Popular Genres: Perform this step using the knowledge gained while performing previous steps.

Your task: Find popular genres

Using consolidate option find the average gross value with respect to their genres

IMDB_Movies (version 1) [Autosaved] - Excel (Product Activation Failed)

Consolidate

Function: Average

Reference: Sheet13!\$A:\$B

All references: Sheet13!\$A:\$B

Use labels in: ☒ Top row ☒ Left column ☐ Create links to source data

OK Close

genres	gross
Action Adventure Fantasy Sci-Fi	423032628
Action Adventure Fantasy	89289910
Action Adventure Thriller	291021565
Action Thriller	141614023
Action Adventure Sci-Fi	623279547
Action Adventure Romance	241063875
Adventure Animation Comedy Family Fantasy Musical Romance	179020854
Action Adventure Sci-Fi	255108370
Action Adventure	262030663

Sort from the largest to the smallest

IMDB_Movies (version 1) [Autosaved] - Excel (Product Activation Failed)

Sort & Filter

Sort Largest to Smallest

Sort Smallest to Largest

Sort Largest to Smallest

Sort Smallest to Largest

genres	gross	genres	gross
Action Adventure Fantasy Sci-Fi	760505847	Action Adventure Fantasy Sci-Fi	296684758.3
Action Adventure Fantasy	309404152	Action Adventure Fantasy	134590736.2
Action Adventure Thriller	200074175	Action Adventure Thriller	102495294.2
Action Thriller	448130642	Action Thriller	73958998.09
Action Adventure Sci-Fi	73058679	Action Adventure Sci-Fi	193550191.5
Action Adventure Romance	336530303	Action Adventure Romance	155302742.2
Adventure Animation Comedy Family Fantasy Musical Romance	200807262	Adventure Animation Comedy Family Fantasy Musical Romance	209078740.5
Action Adventure Sci-Fi	458991599	Adventure Family Fantasy Mystery	279056317
Adventure Family Fantasy Mystery	301956980	Action Adventure	138869176.3
Action Adventure Sci-Fi	330249062	Action Adventure Western	52483256.75
Action Adventure Sci-Fi	200069408	Action Adventure Family Fantasy	90106054
Action Adventure	168368427	Action Adventure Comedy Family Fantasy Sci-Fi	88384658.75
Action Adventure Fantasy	423032628	Adventure Fantasy	93633631.86
Action Adventure Western	89289910	Action Adventure Drama History	52646074
Action Adventure Fantasy Sci-Fi	291021565	Adventure Family Fantasy	133455583.2
Action Adventure Family Fantasy	141614023	Action Adventure Drama Romance	154114075.1
Action Adventure Sci-Fi	623279547	Drama Romance	32388924.63
Action Adventure Fantasy	241063875	Action Adventure Sci-Fi Thriller	108151566.4
Action Adventure Comedy Family Fantasy Sci-Fi	179020854	Action Adventure Fantasy Romance	289279970.3
Adventure Fantasy	255108370	Action Adventure Fantasy Sci-Fi Thriller	95709585.08
Action Adventure Fantasy	262030663	Adventure Animation Comedy Family Fantasy	143316248.9

rank	genres	gross
1	Family Sci-Fi	434949459
2	Adventure Animation Drama Family Musical	422783777
3	Adventure Animation Comedy Drama Family Fantasy	356454367
4	Action Biography Drama History Thriller War	350123553
5	Action Adventure Fantasy Sci-Fi	296684758.3
6	Adventure Drama Fantasy Romance	295436148.3
7	Action Adventure Fantasy Romance	289279970.3
8	Adventure Sci-Fi	281666058
9	Adventure Family Fantasy Mystery	279056317
10	Action Adventure Animation Family	261437578
11	Action Adventure Comedy Family Fantasy	250863268
12	Adventure Comedy Family Mystery Sci-Fi	250147615
13	Animation Comedy Family Sci-Fi	246459955
14	Adventure Animation Comedy Family Fantasy Romance	243310827.5
15	Action Adventure Family Fantasy Romance	241407328
16	Adventure Drama Fantasy	221186651.3
17	Adventure Sci-Fi Thriller	221050858.6
18	Adventure Animation Comedy Family Fantasy Romance	220068175
19	Animation Comedy Family Fantasy Music	218469863.5
20	Drama Fantasy Romance Thriller	217631306
21	Action Adventure Drama Fantasy	212116144.5

rank them using the auto-fill option

Family|sci-fi has the highest mean

- 6) **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.

use nested if function to get the data of 3 actors separately

=if(a1:a5737, "Brad Pitt", "accept", if(a1:a5737, "Leonardo DiCaprio", "accept", if(a1:a5737, "Meryl Streep", "accept", "reject"))))

	A	B	C	D	E
1	actor_1_name	movie_title	num_critic_for_reviews	num_user_for_reviews	
2	CCH Pounder	Avatar	723	3054	=IF(A1:A5737="Brad Pitt","accept",IF(A1:A5737="Leonardo DiCaprio",
3	Johnny Depp	Pirates of the Caribbean: At World's End	302	1238	"accept",IF(A1:A5737="Meryl Streep","accept","reject"))
4	Christoph Waltz	Spectre	602	994	ri IF(logical_test,[value_if_true],[value_if_false])
5	Tom Hardy	The Dark Knight Rises	813	2701	reject
6	Daryl Sabara	John Carter	462	738	reject
7	J.K. Simmons	Spider-Man 3	392	1902	reject
8	Brad Garrett	Tangled	324	387	reject
9	Chris Hemsworth	Avengers: Age of Ultron	635	1117	reject
10	Alan Rickman	Harry Potter and the Half-Blood Prince	375	973	reject
11	Henry Cavill	Batman v Superman: Dawn of Justice	673	3018	reject
12	Kevin Spacey	Superman Returns	434	2367	reject
13	Giancarlo Giannini	Quantum of Solace	403	1243	reject
14	Johnny Depp	Pirates of the Caribbean: Dead Man's Chest	313	1832	reject
15	Johnny Depp	The Lone Ranger	450	711	reject
16	Henry Cavill	Man of Steel	733	2536	reject
17	Peter Dinklage	The Chronicles of Narnia: Prince Caspian	258	438	reject
18	Chris Hemsworth	The Avengers	703	1722	reject
19	Johnny Depp	Pirates of the Caribbean: On Stranger Tides	448	484	reject
20	Will Smith	Men in Black 3	451	341	reject
21	Aidan Turner	The Hobbit: The Battle of the Five Armies	422	802	reject
22	Emma Stone	The Amazing Spider-Man	599	1225	reject

Delete the rejected rows by selecting them using ctrl + f using find&replace dialoguebox

Find and Replace

Find what: reject

Within: Sheet

Search: By Rows

Look in: Values

Options

Find All Find Next Close

Book Sheet Name Cell Value Formula

IMDB... Sheet5 SES... reject =IF(A3795="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3796="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3797="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3798="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3799="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3800="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3801="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3802="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3803="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3804="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3805="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3806="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3807="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3808="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3809="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3810="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3811="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3812="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3813="Meryl Streep","accept"...

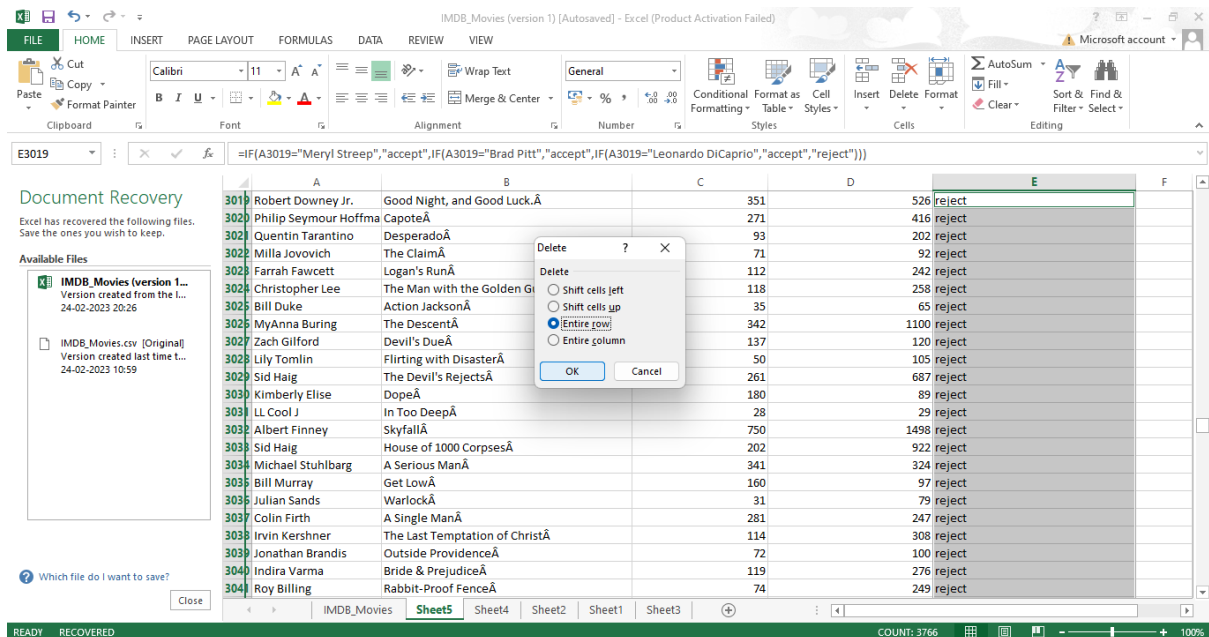
IMDB... Sheet5 SES... reject =IF(A3814="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3815="Meryl Streep","accept"...

IMDB... Sheet5 SES... reject =IF(A3816="Meryl Streep","accept"...

READY RECOVERED COUNT: 3766 100%

Delete the selected row using shortcut ctrl+ - -> entire row



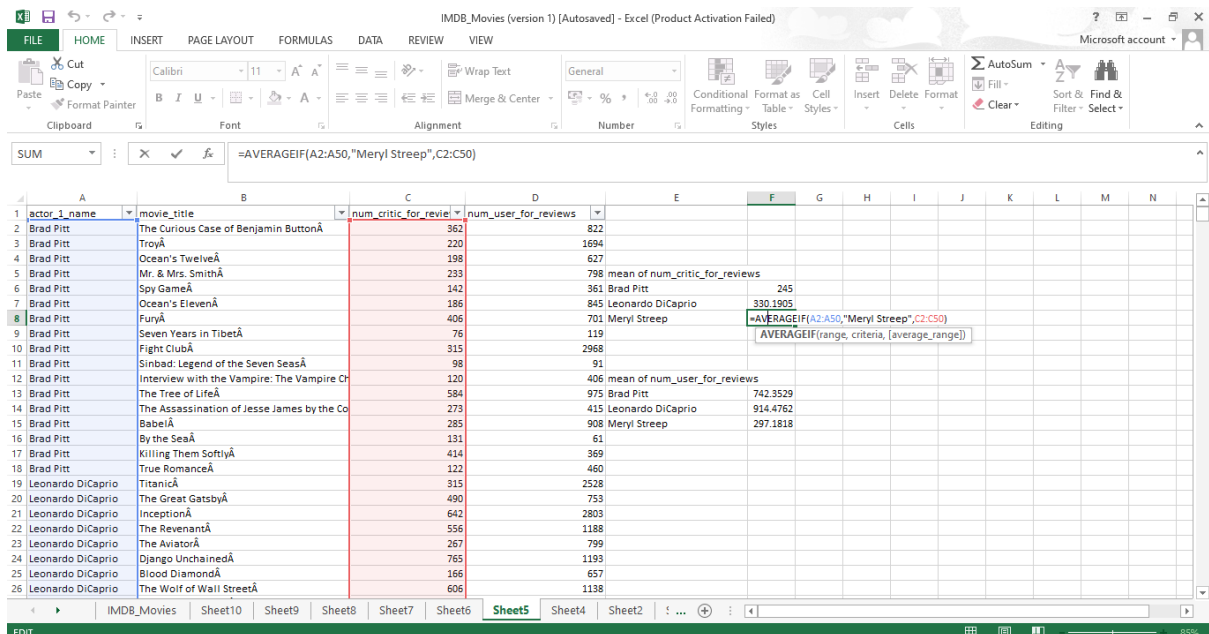
To find the mean of the 3 actors use average if condition

mean of num_critic_for_reviews

=AVERAGEIF(A2:A50,"Meryl Streep",C2:C50)

=AVERAGEIF(A2:A50,"Leonardo DiCaprio",C2:C50)

=AVERAGEIF(A2:A50,"Brad Pitt",C2:C50)



mean of num_user_for_reviews

=AVERAGEIF(A2:A50,"Brad Pitt",D2:D50)

=AVERAGEIF(A2:A50,"Leonardo DiCaprio",D2:D50)

=AVERAGEIF(A2:A50,"Meryl Streep",D2:D50)

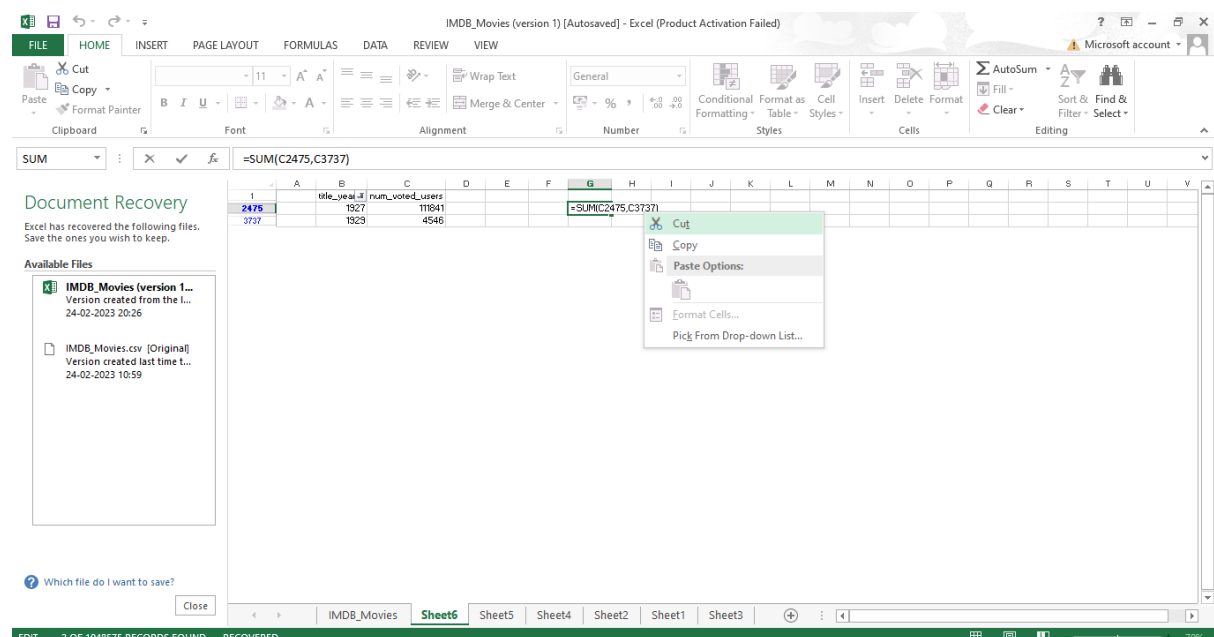
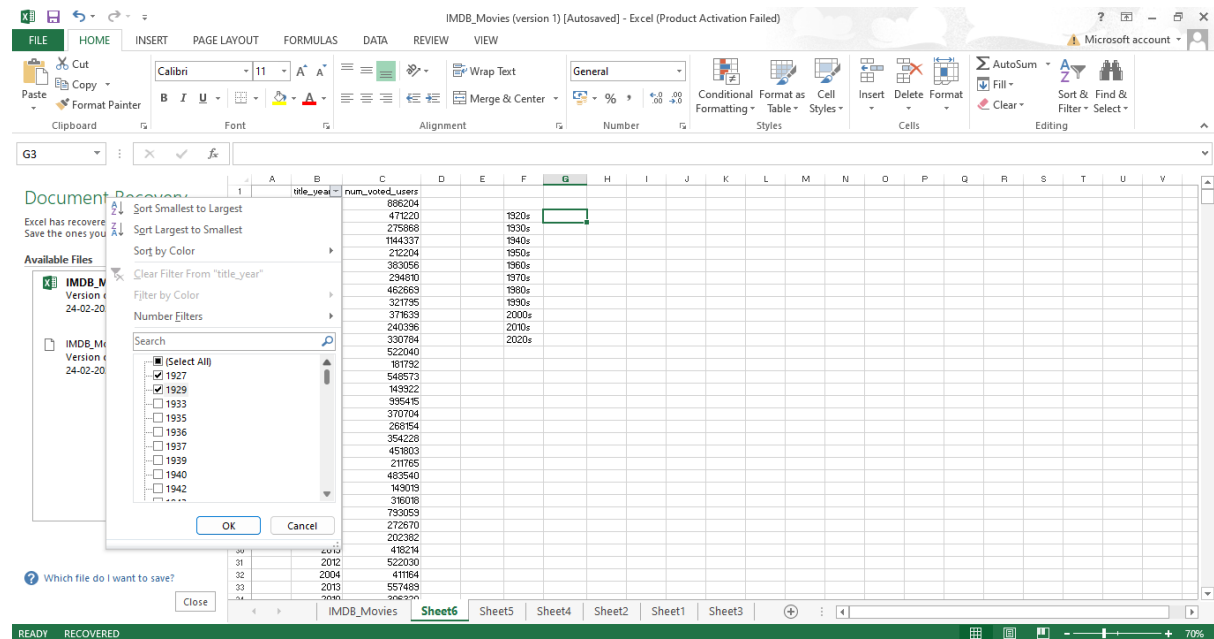
actor_1_name	movie_title	num_critics_for_review	num_user_for_reviews
Brad Pitt	The Curious Case of Benjamin Button	362	822
Brad Pitt	Troy	220	1694
Brad Pitt	Ocean's Twelve	198	627
Brad Pitt	Mr. & Mrs. Smith	233	798
Brad Pitt	Spy Game	142	361
Brad Pitt	Ocean's Eleven	186	845
Brad Pitt	Fury	406	701
Brad Pitt	Seven Years in Tibet	76	119
Brad Pitt	Fight Club	315	2968
Brad Pitt	Sinbad: Legend of the Seven Seas	98	91
Brad Pitt	Interview with the Vampire: The Vampire Chronicles	120	406
Brad Pitt	The Tree of Life	584	975
Brad Pitt	The Assassination of Jesse James by the Coward Jim Bowie	273	415
Brad Pitt	Babel	285	908
Brad Pitt	By the Sea	131	61
Brad Pitt	Killing Them Softly	414	369
Brad Pitt	True Romance	122	460
Leonardo DiCaprio	Titanic	315	2528
Leonardo DiCaprio	The Great Gatsby	490	753
Leonardo DiCaprio	Inception	642	2803
Leonardo DiCaprio	The Revenant	556	1188
Leonardo DiCaprio	The Aviator	267	799
Leonardo DiCaprio	Django Unchained	765	1193
Leonardo DiCaprio	Blood Diamond	166	657
Leonardo DiCaprio	The Wolf of Wall Street	606	1138

actor_1_name	movie_title	num_critics_for_review	num_user_for_reviews
Brad Pitt	The Curious Case of Benjamin Button	362	822
Brad Pitt	Troy	220	1694
Brad Pitt	Ocean's Twelve	198	627
Brad Pitt	Mr. & Mrs. Smith	233	798
Brad Pitt	Spy Game	142	361
Brad Pitt	Ocean's Eleven	186	845
Brad Pitt	Fury	406	701
Brad Pitt	Seven Years in Tibet	76	119
Brad Pitt	Fight Club	315	2968
Brad Pitt	Sinbad: Legend of the Seven Seas	98	91
Brad Pitt	Interview with the Vampire: The Vampire Chronicles	120	406
Brad Pitt	The Tree of Life	584	975
Brad Pitt	The Assassination of Jesse James by the Coward Jim Bowie	273	415
Brad Pitt	Babel	285	908
Brad Pitt	By the Sea	131	61
Brad Pitt	Killing Them Softly	414	369
Brad Pitt	True Romance	122	460
Leonardo DiCaprio	Titanic	315	2528
Leonardo DiCaprio	The Great Gatsby	490	753
Leonardo DiCaprio	Inception	642	2803
Leonardo DiCaprio	The Revenant	556	1188
Leonardo DiCaprio	The Aviator	267	799
Leonardo DiCaprio	Django Unchained	765	1193
Leonardo DiCaprio	Blood Diamond	166	657
Leonardo DiCaprio	The Wolf of Wall Street	606	1138

the actors which have the highest mean is Leonardo DiCaprio

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 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.

For calculating the sum of votes in each decade use the filter option and select the required rows to which you want to perform the sum operation



If there are large number of values to perform the sum function use the sum if function

`=SUM(SUMIF(B2:B3168,{"1990","1991","1992","1993","1994","1995","1996","1997","1998","1999"},C2:C3758))`

To get the sum of the certain values

IMDB_Movies (version 1) [Autosaved] - Excel (Product Activation Failed)

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW

Clipboard Font Alignment Number Styles Cells Editing

SUM =SUM(SUMIF(B2:B3168,{"1990","1991","1992","1993","1994","1995","1996","1997","1998","1999"},C2:C3758))

Document Recovery

Excel has recovered the following files. Save the ones you wish to keep.

Available Files

- IMDB_Movies (version 1) ... Version created from the l... 24-02-2023 20:26
- IMDB_Movies.csv [Original] Version created last time t... 24-02-2023 10:59

Which file do I want to save?

Close

	A	B	C	D	E	F	G	H	I	J	K	L
1		title_year	num_voted_users									
2		2009	886204									
3		2007	471220			1920s	116387					
4		2015	275868			1930s	804839					
5		2012	1144337			1940s	636002					
6		2012	212204			1950s	678336					
7		2007	383056			1960s	3120472					
8		2010	294810			1970s	8854391					
9		2015	462669			1980s	20092021					
10		2009	321795			1990s	=SUM(SUMIF(B2:B3168,{"1990","1991","1992","1993","1994","1995","1996","1997","1998","1999"},C2:C3758))					
11		2016	371639			2000s						
12		2006	240396			2010s	SUMIF(range, criteria, [sum_range])					
13		2008	330784									
14		2006	522040									
15		2013	181792									
16		2013	548573									
17		2008	149922									
18		2012	995415									
19		2011	370704									
20		2012	268154									
21		2014	354228									
22		2012	451803									

IMDB_Movies Sheet7 Sheet6 Sheet5 Sheet4 Sheet2 Sheet1

EDIT RECOVERED 100%

IMDB_Movies (version 1) [Autosaved] - Excel (Product Activation Failed)

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW CHART TOOLS DESIGN FORMAT

Clipboard Font Alignment Number Styles Cells Editing

Chart 2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		title_year	num_voted_users												
2		2009	886204			decade	df_by_decade								
3		2007	471220			1920s	116387								
4		2015	275868			1930s	804839								
5		2012	1144337			1940s	636002								
6		2012	212204			1950s	678336								
7		2007	383056			1960s	3120472								
8		2010	294810			1970s	8854391								
9		2015	462669			1980s	20092021								
10		2009	321795			1990s	65124301								
11		2016	371639			2000s	159944165								
12		2006	240396			2010s	112974368								
13		2008	330784												
14		2006	522040												
15		2013	181792												
16		2013	548573												
17		2008	149922												
18		2012	995415												
19		2011	370704												
20		2012	268154												
21		2014	354228												

df_by_decade

Sheet12 Sheet11 Sheet10 Sheet13 Sheet9 Sheet8 Sheet7 Sheet6 Sheet5 Sher ...

READY 100%

