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.

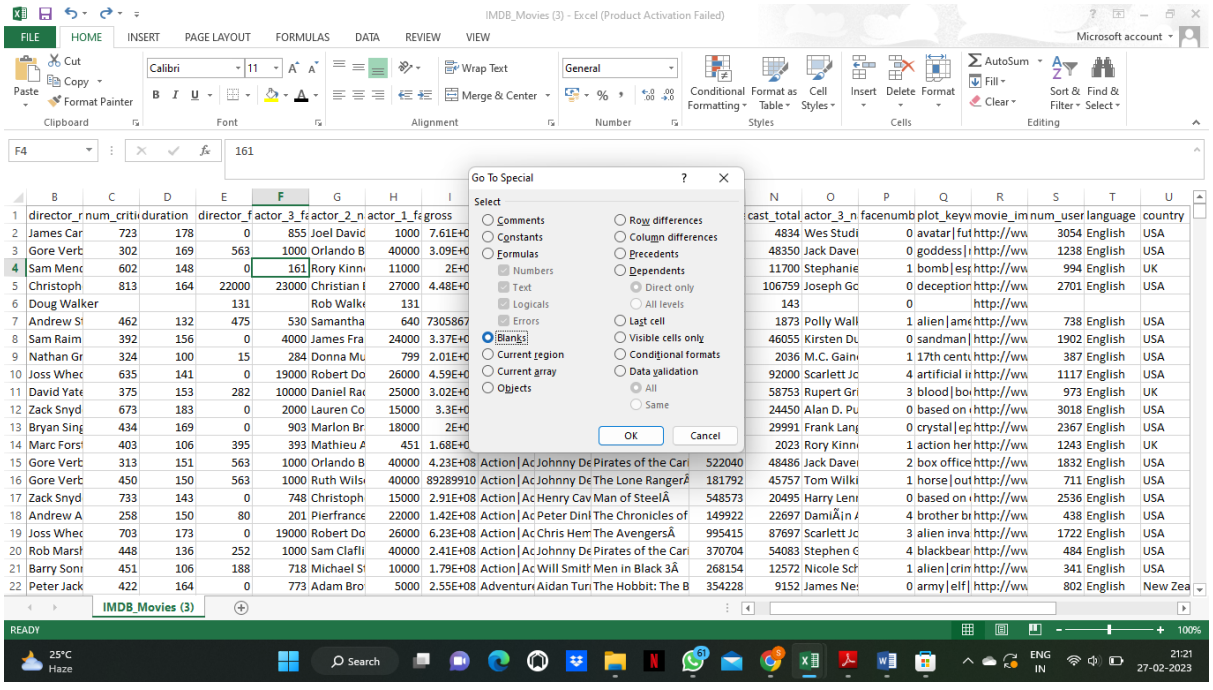
RESULT:

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

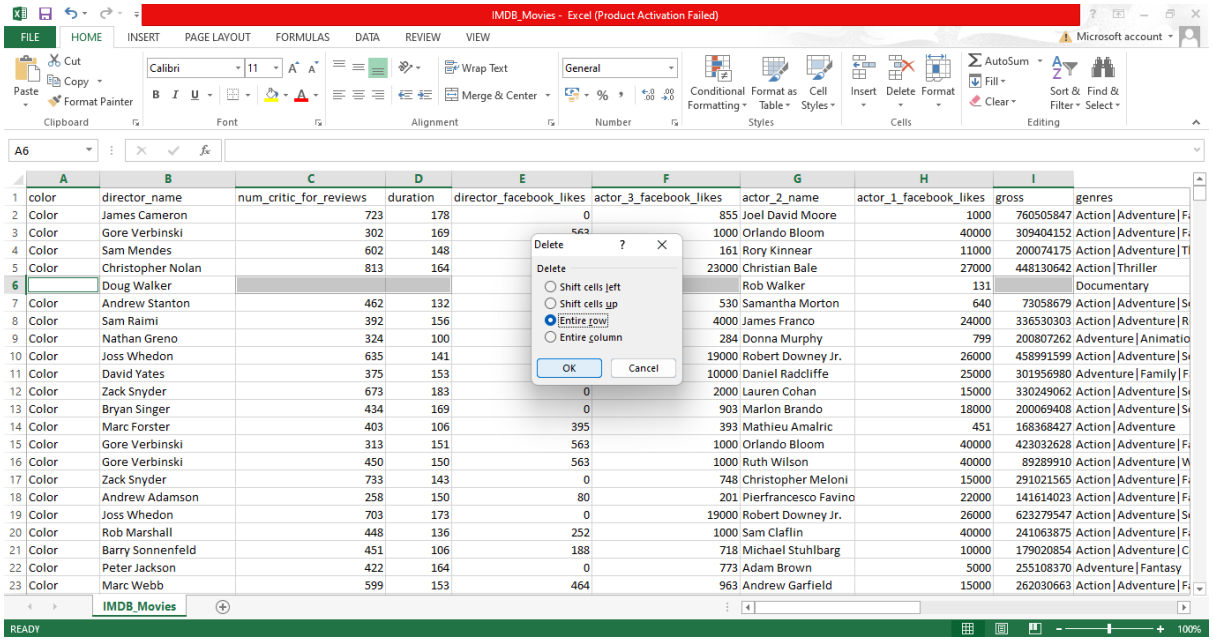
1) Cleaning the data:: PThis is one of the most important step to perform before moving forward with the analysis. Use your knowledge learned till now to do this. (Dropping columns, removing null values, etc.)
Your task: Clean the data

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

To select all the black cells



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



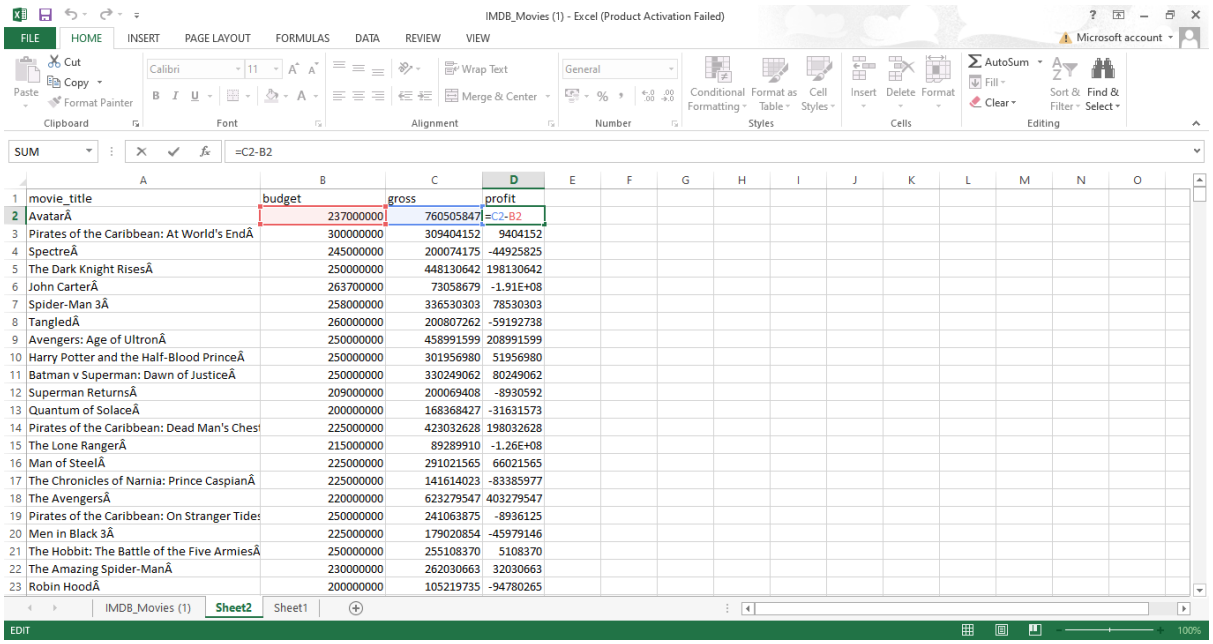
2)Movies with highest profit: Create a new column called profit which contains the difference of the two columns: gross and budget. Sort the column using the profit column as reference. Plot profit (y-axis) vs budget (x- axis) and observe the outliers using the appropriate chart type.
Your task: Find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

Extrate the data from the given database

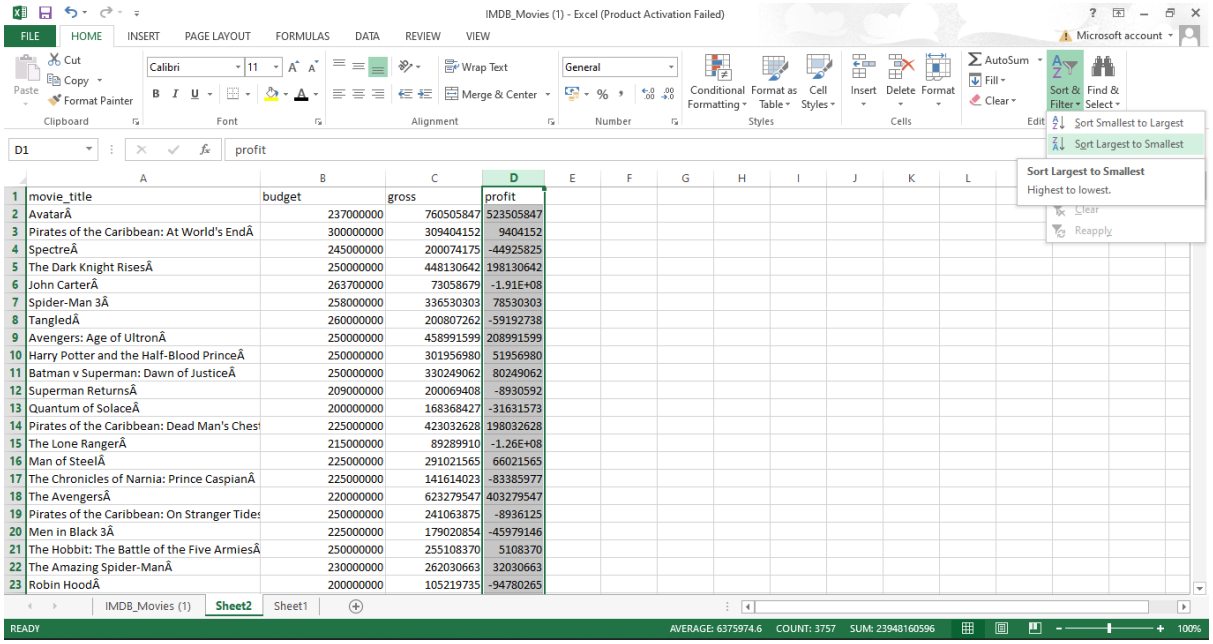
Profit=gross-budget

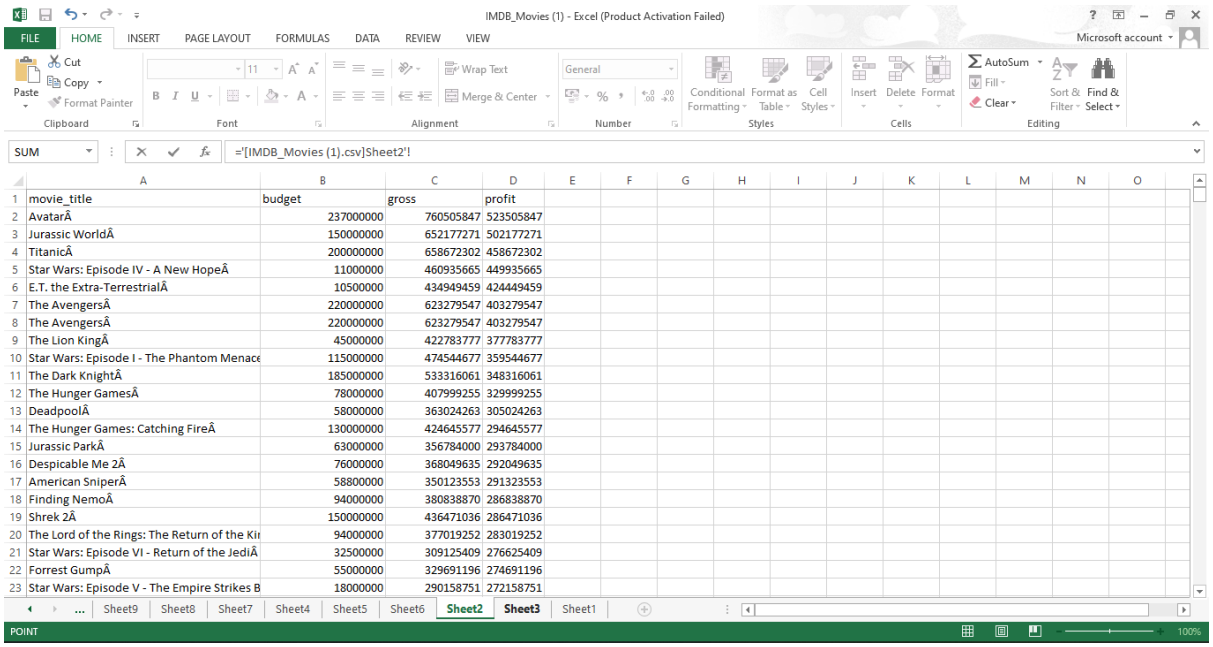
=c2-b2

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

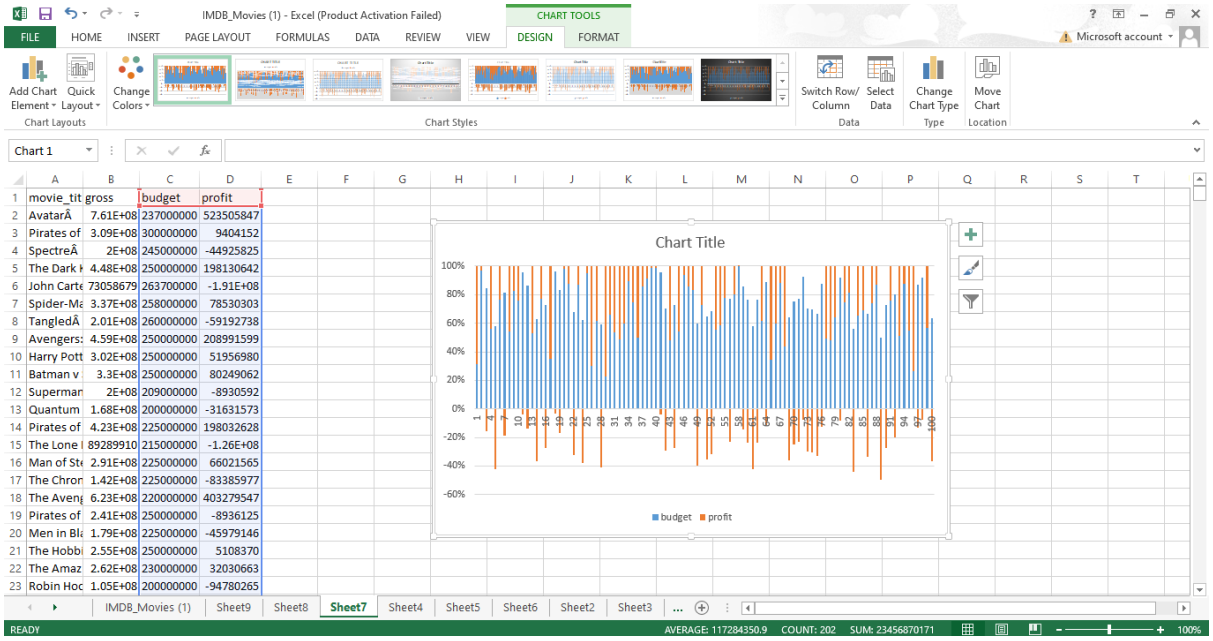


Sort the data from largest to smallest to get the movie with the highest profit





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

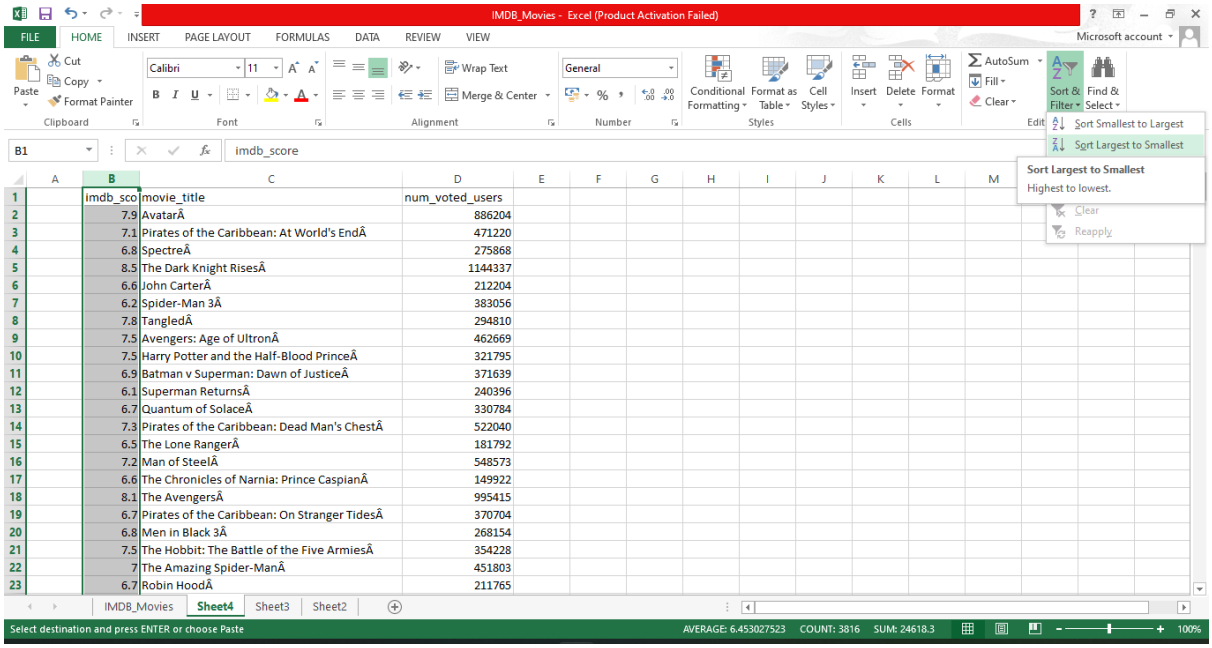


	A	B	C	D
1	movie_tit	gross	budget	profit
2	AvatarÂ	7.61E+08	237000000	523505847
3	Pirates of	3.09E+08	300000000	9404152
4	SpectreÃ	2E+08	245000000	-44925825
5	The Dark I	4.48E+08	250000000	198130642
6	John Carte	73058679	263700000	-1.91E+08
7	Spider-Ma	3.37E+08	258000000	78530303
8	TangledÃ	2.01E+08	260000000	-59192738
9	Avengers:	4.59E+08	250000000	208991599
10	Harry Pott	3.02E+08	250000000	51956980
11	Batman v	3.3E+08	250000000	80249062
12	Superman	2E+08	209000000	-8930592
13	Quantum	1.68E+08	200000000	-31631573
14	Pirates of	4.23E+08	225000000	198032628
15	The Lone I	89289910	215000000	-1.26E+08
16	Man of Ste	2.91E+08	225000000	66021565
17	The Chron	1.42E+08	225000000	-83385977
18	The Avenş	6.23E+08	220000000	403279547
19	Pirates of	2.41E+08	250000000	-8936125
20	Men in Bla	1.79E+08	225000000	-45979146
21	The Hobbi	2.55E+08	250000000	5108370
22	The Amaz	2.62E+08	230000000	32030663
23	Robin Hoc	1.05E+08	200000000	-94780265

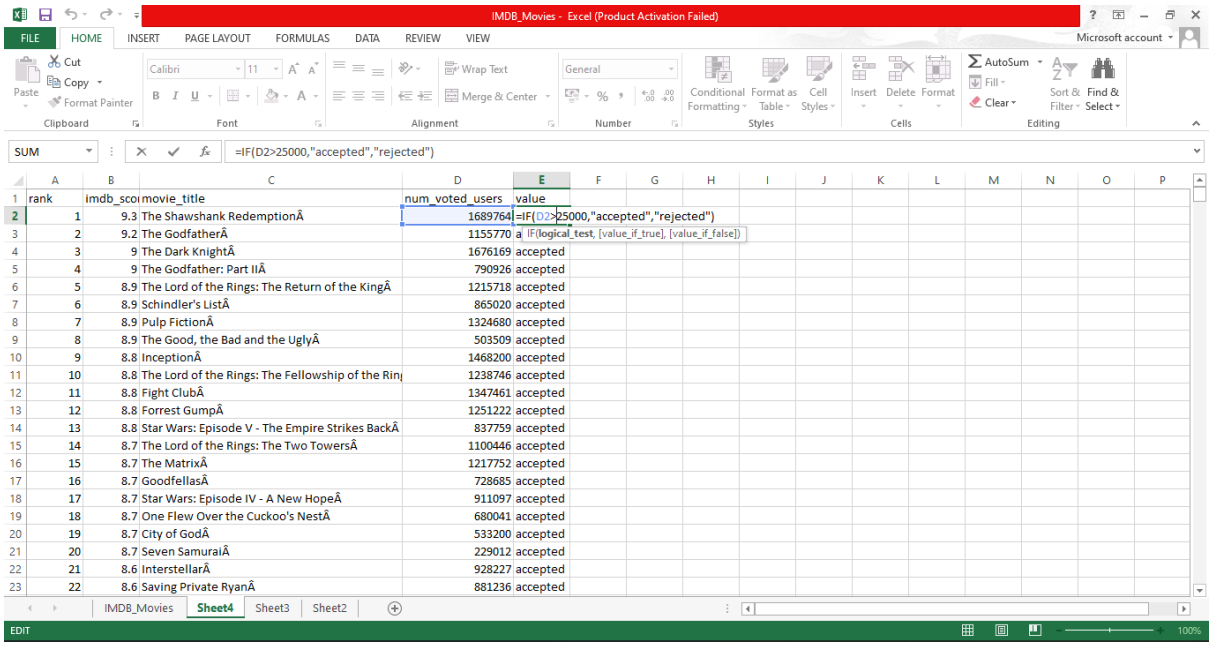
3) **Top 250:** Create a new column IMDb_Top_250 and store the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb_score). Also make sure that for all of these movies, the num_voted_users is greater than 25,000. Also add a Rank column containing the values 1 to 250 indicating 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 ibm score from largest to the smallest so that it becomes easy in performing the calculation

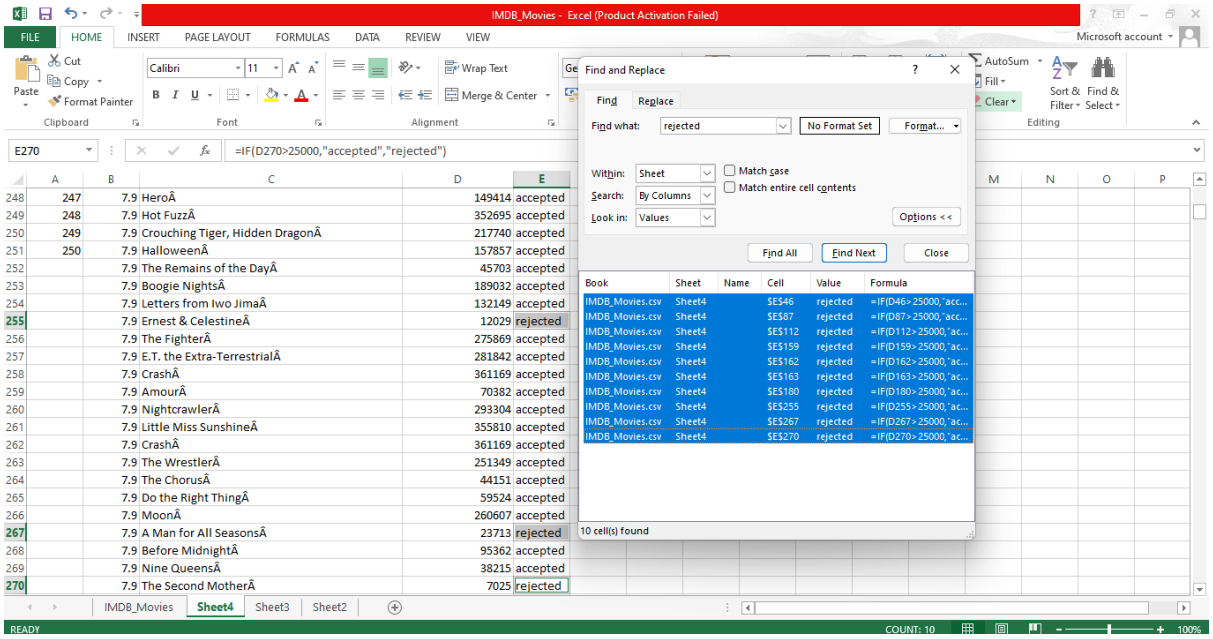


Creating a new column value and accepting the rows that have num_voted_users >25000

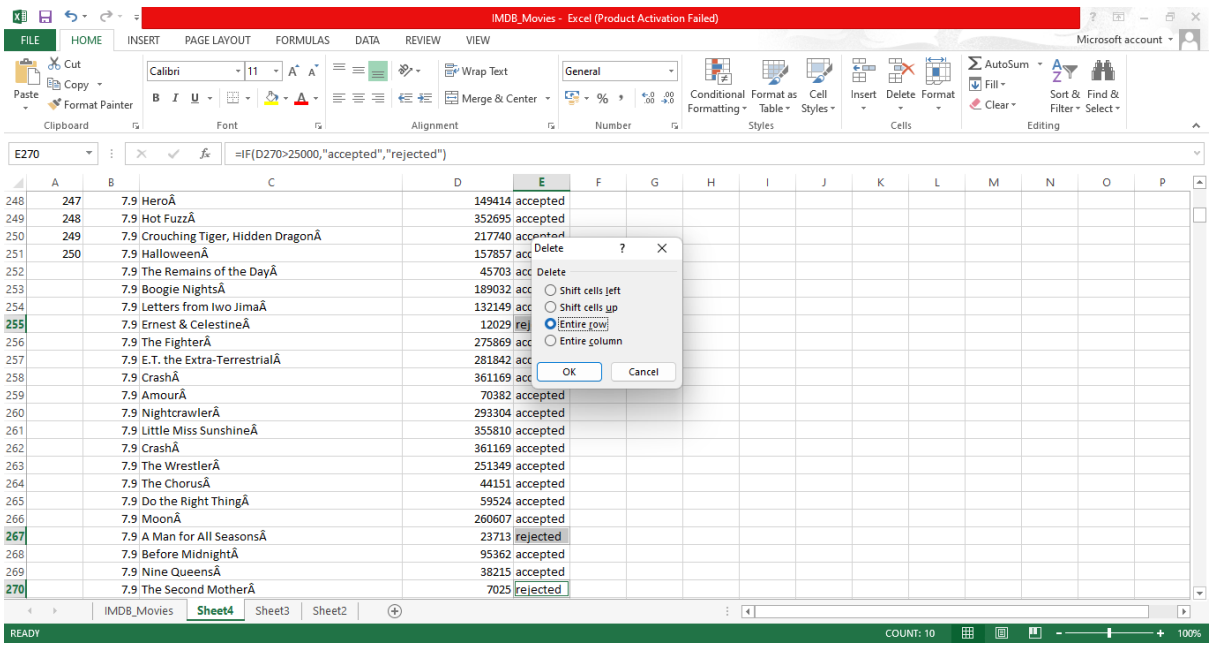


Selecting the cells with value “rejected”

By Using ctrl+f and find option

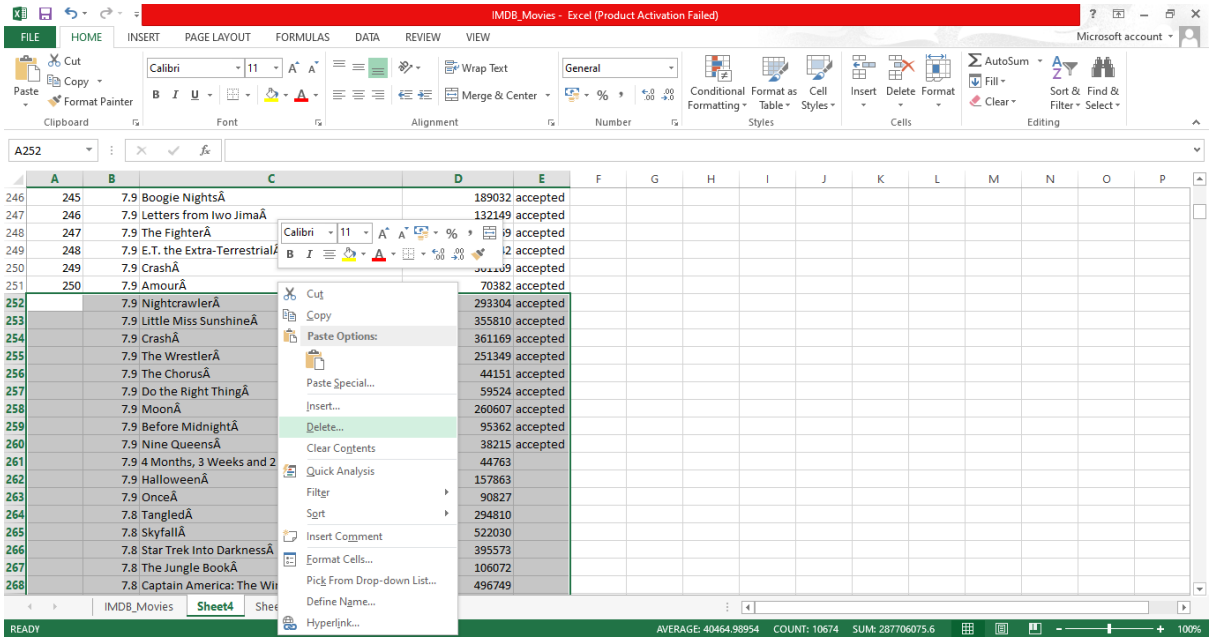


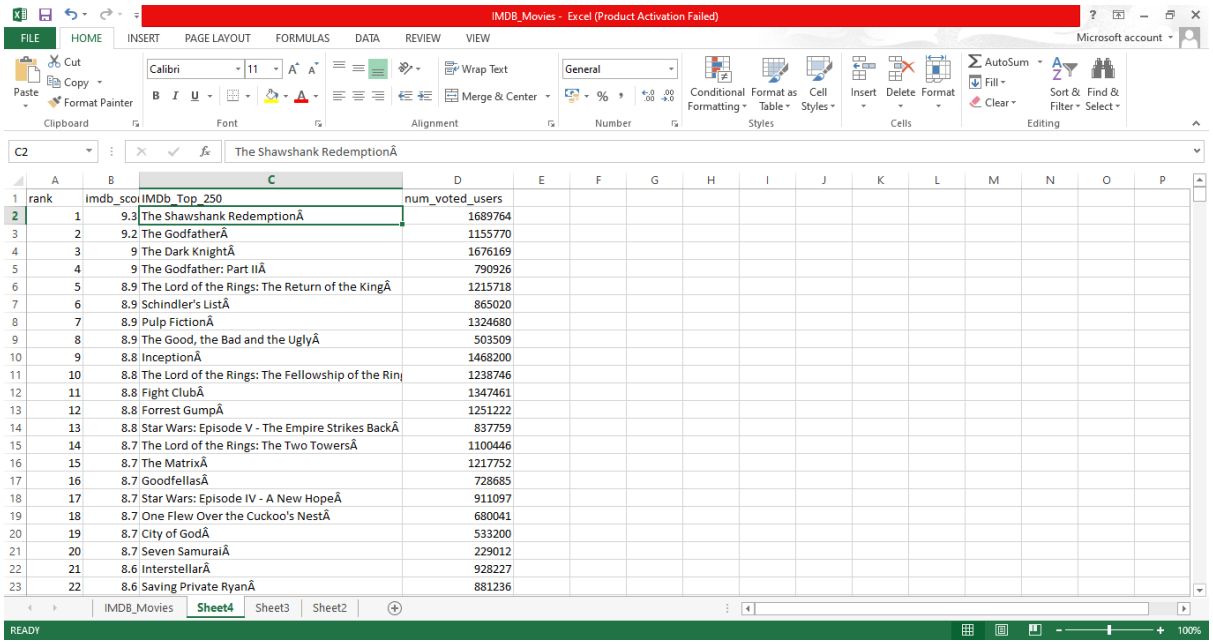
Deleting the selected rows by ctrl+ - and selecting entire row radio button



Ranking them from 1-250 by autofill option

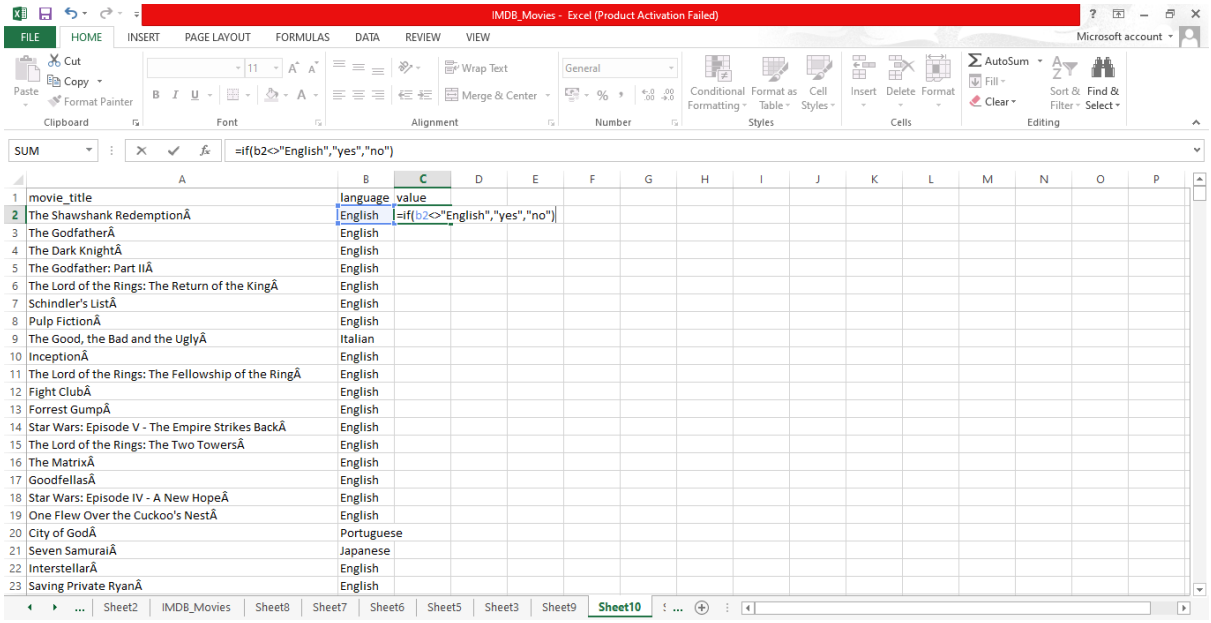
And deleting the excess data by seleting them with the range option





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

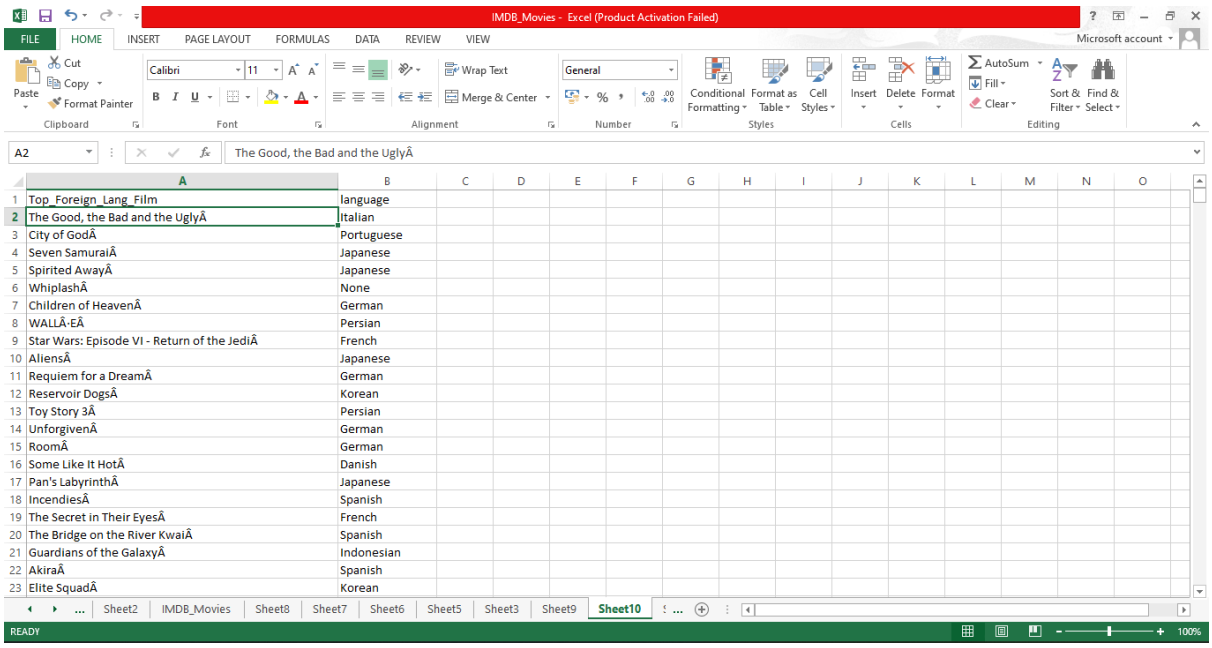
=if(<>”English,”yes”,”no”)



The rows with English as its language are selected

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

And deleting them using ctrl+ - command

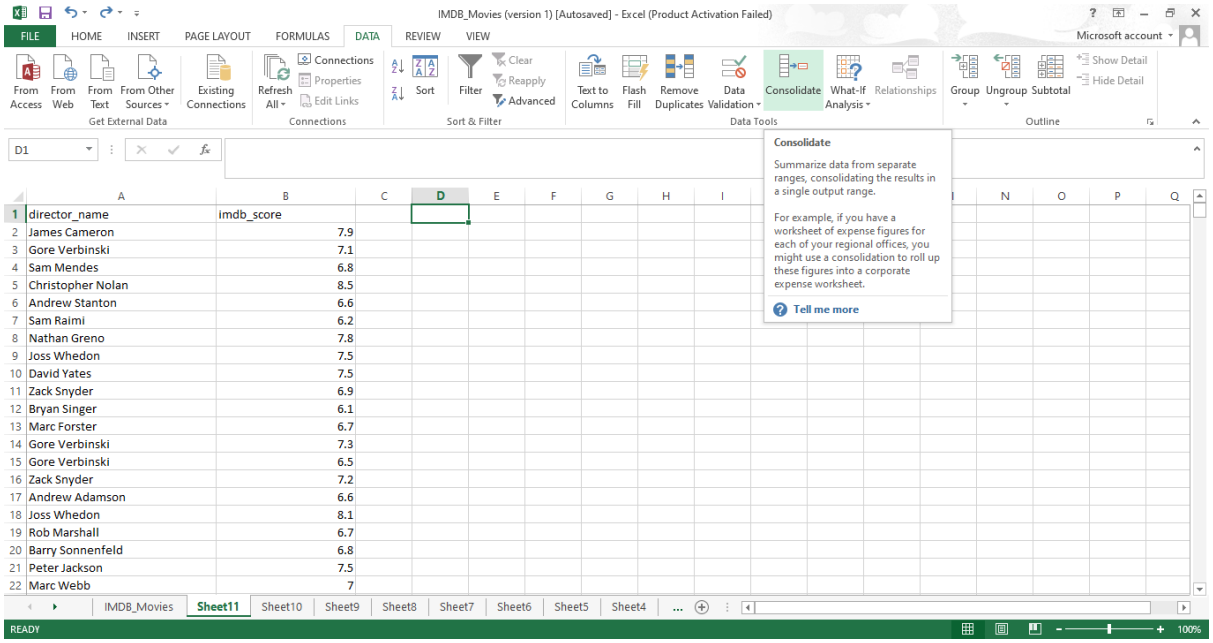


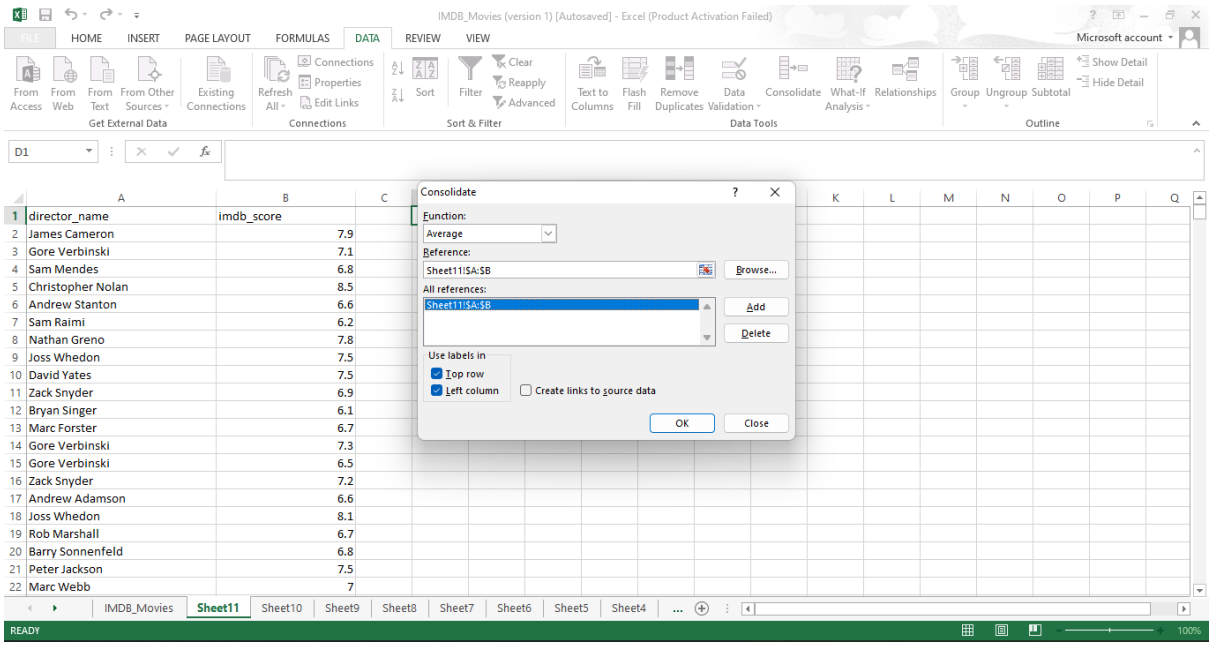
4) **Best Directors:** TGroup the column using the director_name column.

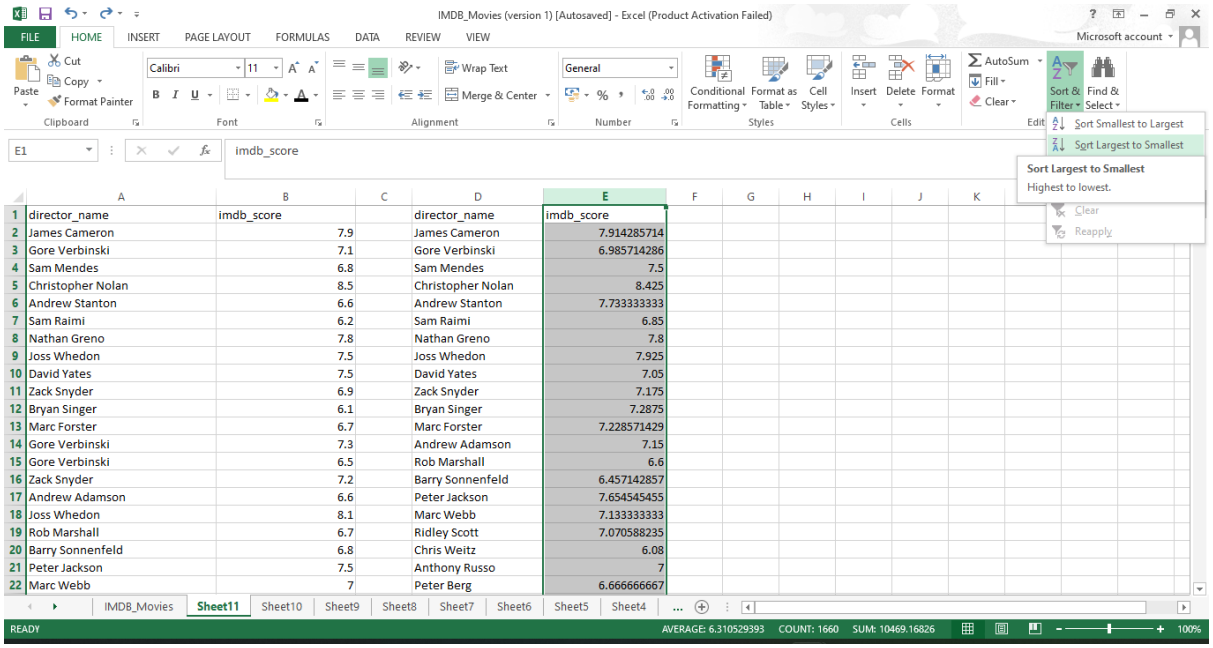
Find out the top 10 directors for whom the mean of imdb_score is the highest and store them in a new column top10director. In case of a tie in IMDb score

between two directors, sort them alphabetically.
Your task: Find the best directors

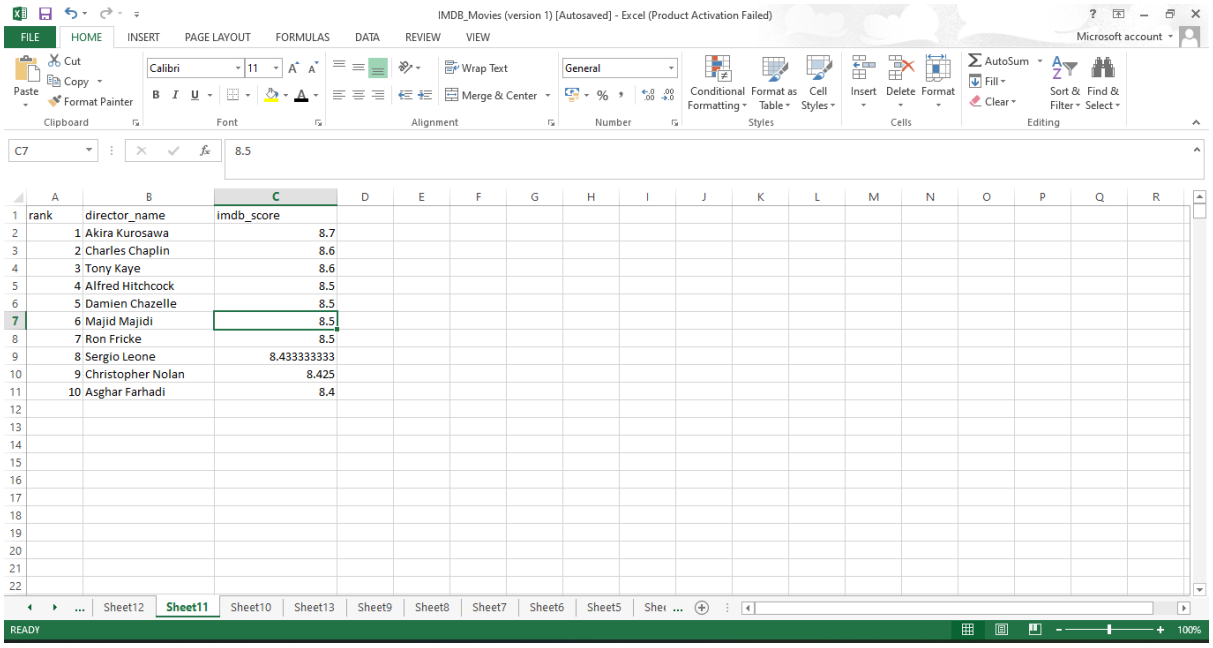
Using consolidate option find the average imbd_score of each director after groping them according to their names



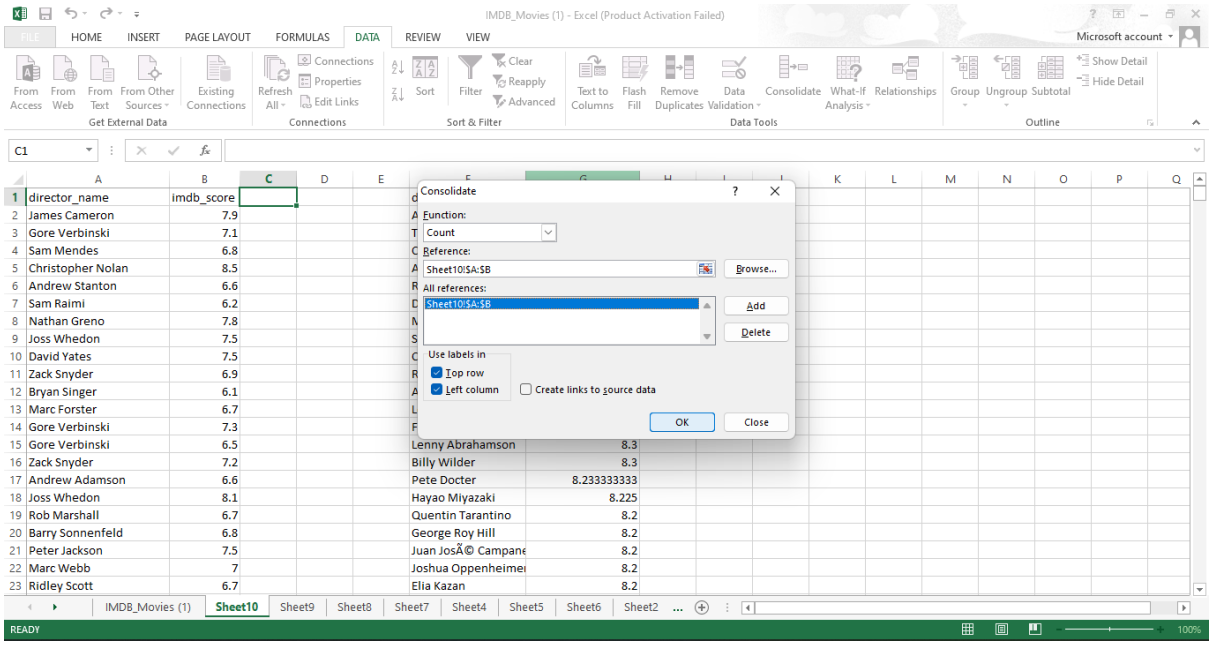


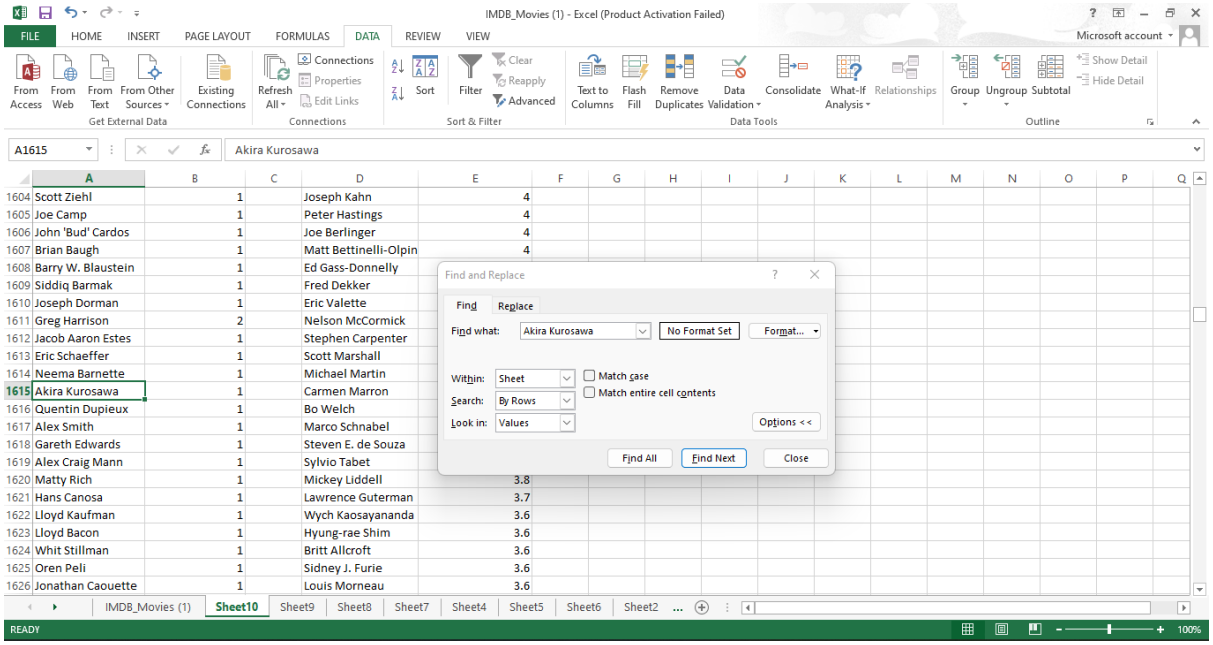


To find the top 10 short it from the highest to lowest



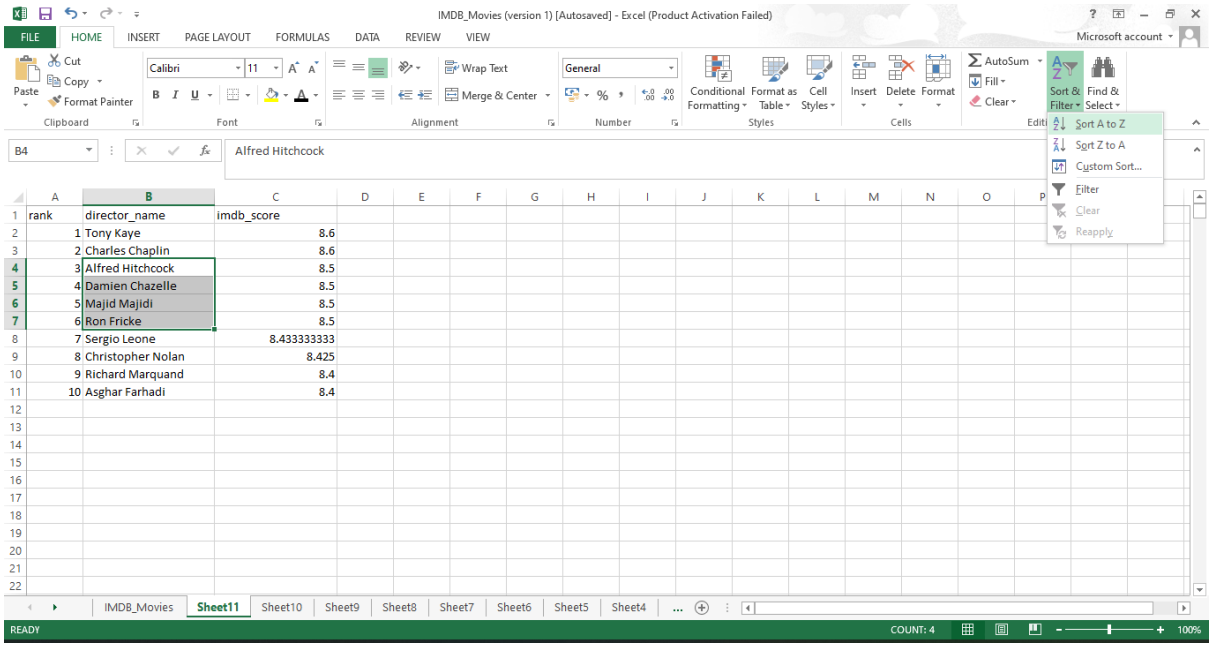
use the count option groping by director names to see whether any of them have done only one film

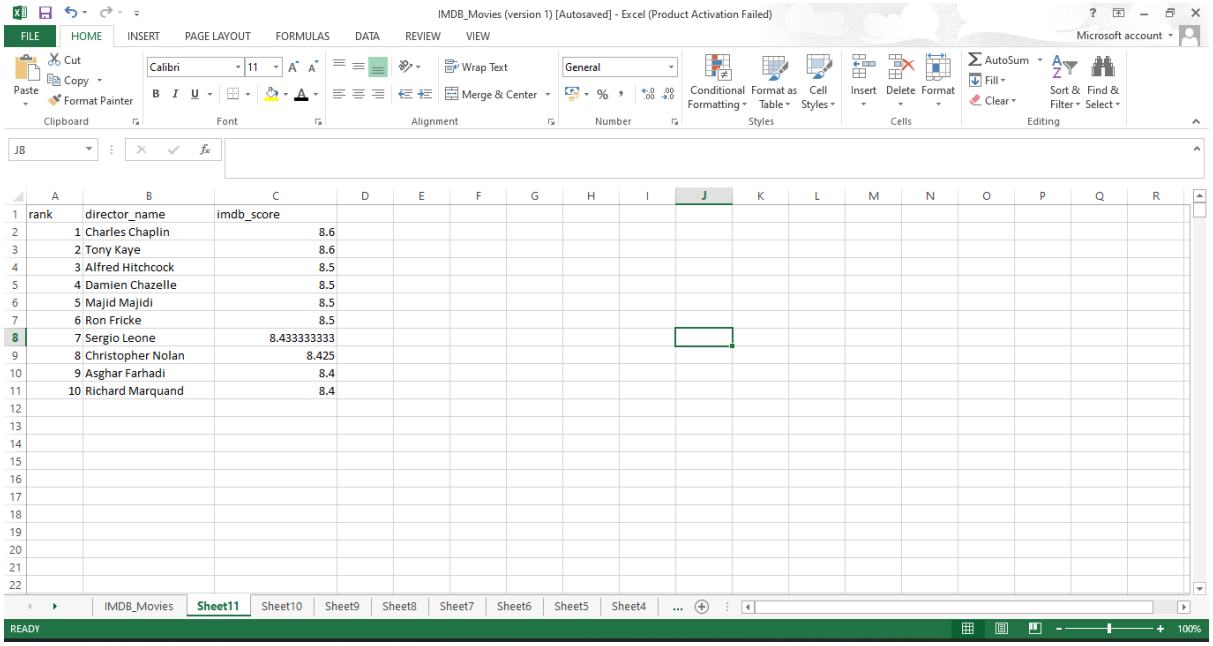




If u want to have only the directors who have done more than one film delete them

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



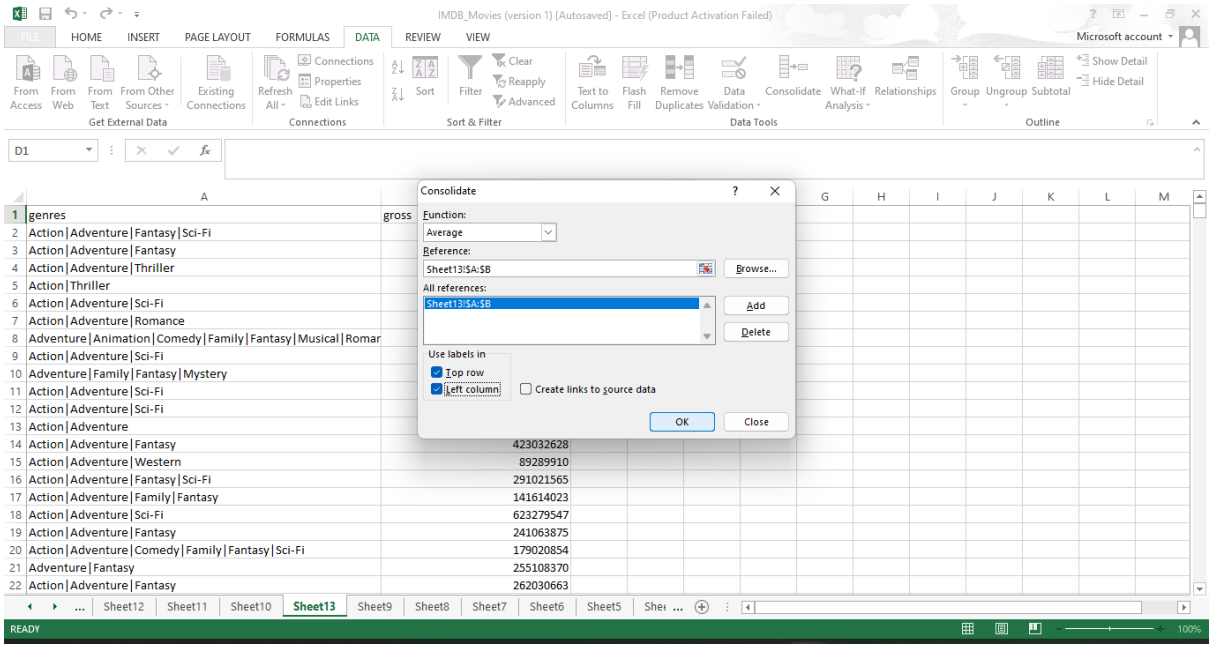


rank	director_name	imdb_score
1	Tony Kaye	8.6
2	Charles Chaplin	8.6
3	Alfred Hitchcock	8.5
4	Damien Chazelle	8.5
5	Majid Majidi	8.5
6	Ron Fricke	8.5
7	Sergio Leone	8.433333333
8	Christopher Nolan	8.425
9	Richard Marquand	8.4
10	Asghar Farhadi	8.4

rank	director_name	imdb_score
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

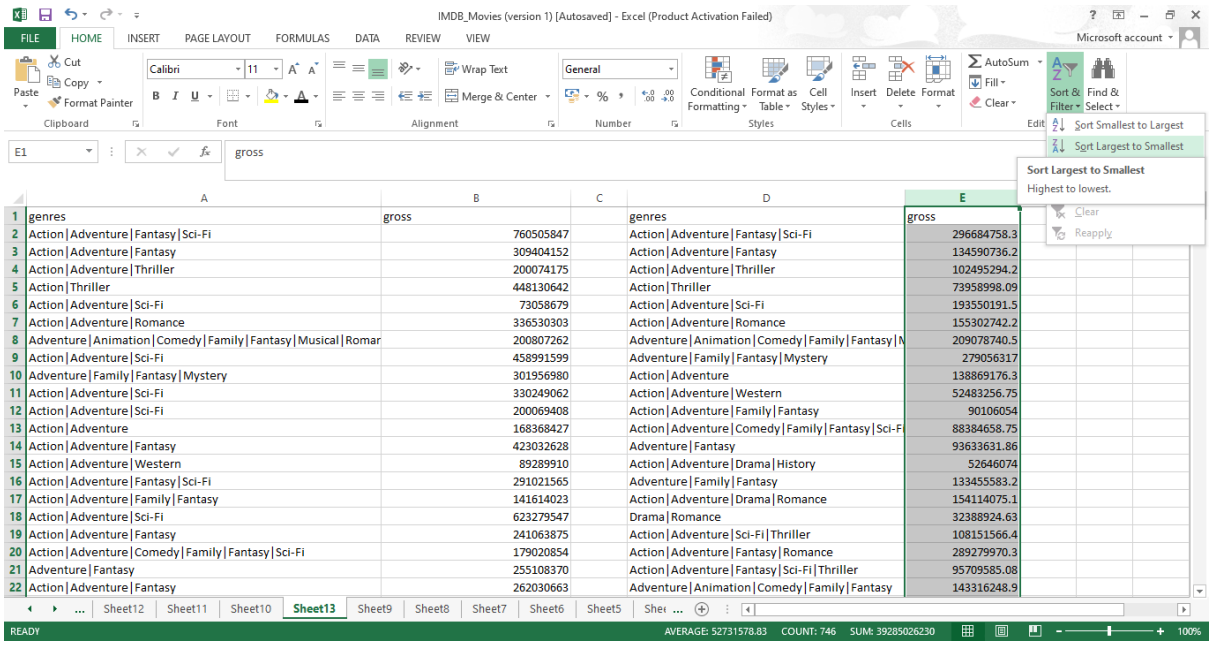
5 **)Popular Genres:** Perform this step using the knowledge gained while performing previous steps.
Your task: Find popular genres

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

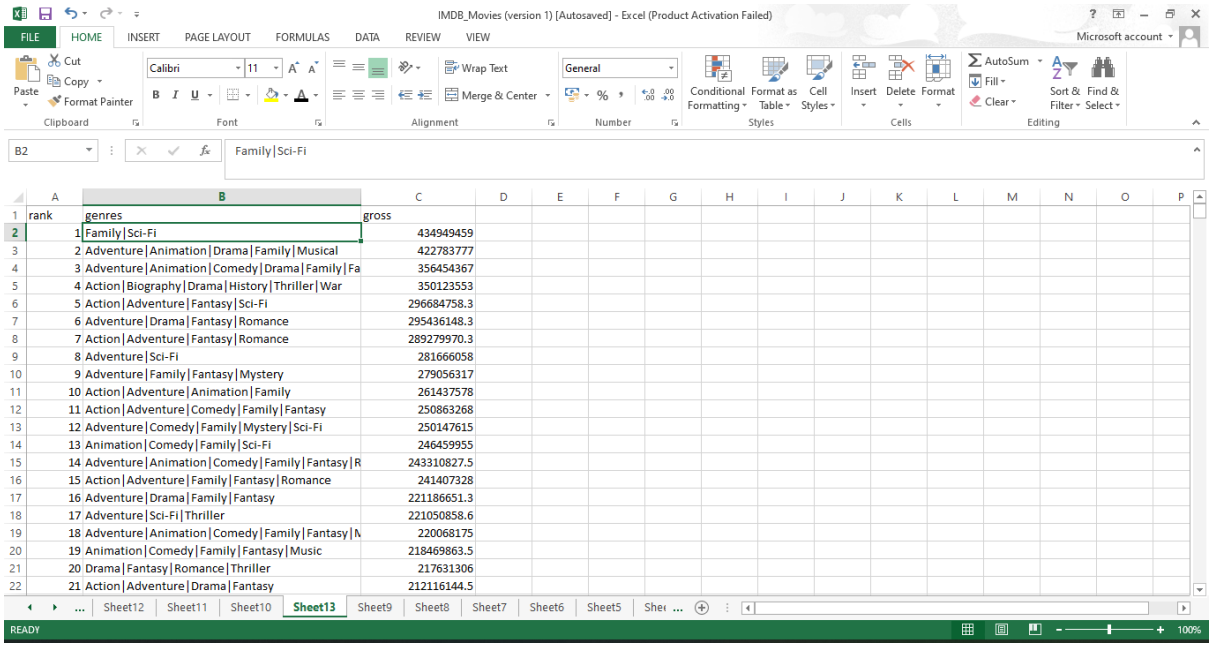


	A	gross
1	genres	
2	Action|Adventure|Fantasy|Sci-Fi	
3	Action|Adventure|Fantasy	
4	Action|Adventure|Thriller	
5	Action|Thriller	
6	Action|Adventure|Sci-Fi	
7	Action|Adventure|Romance	
8	Adventure|Animation|Comedy|Family|Fantasy|Musical|Romar	
9	Action|Adventure|Sci-Fi	
10	Adventure|Family|Fantasy|Mystery	
11	Action|Adventure|Sci-Fi	
12	Action|Adventure|Sci-Fi	
13	Action|Adventure	
14	Action|Adventure|Fantasy	423032628
15	Action|Adventure|Western	89289910
16	Action|Adventure|Fantasy|Sci-Fi	291021565
17	Action|Adventure|Family|Fantasy	141614023
18	Action|Adventure|Sci-Fi	623279547
19	Action|Adventure|Fantasy	241063875
20	Action|Adventure|Comedy|Family|Fantasy|Sci-Fi	179020854
21	Adventure|Fantasy	255108370
22	Action|Adventure|Fantasy	262030663

Sort from the largest to the smallest



	A	B	C	D	E
1	genres	gross		genres	gross
2	Action|Adventure|Fantasy|Sci-Fi	760505847		Action|Adventure|Fantasy|Sci-Fi	296684758.3
3	Action|Adventure|Fantasy	309404152		Action|Adventure|Fantasy	134590736.2
4	Action|Adventure|Thriller	200074175		Action|Adventure|Thriller	102495294.2
5	Action|Thriller	448130642		Action|Thriller	73958998.09
6	Action|Adventure|Sci-Fi	73058679		Action|Adventure|Sci-Fi	193550191.5
7	Action|Adventure|Romance	336530303		Action|Adventure|Romance	155302742.2
8	Adventure|Animation|Comedy|Family|Fantasy|Musical|Romar	200807262		Adventure|Animation|Comedy|Family|Fantasy|M	209078740.5
9	Action|Adventure|Sci-Fi	458991599		Adventure|Family|Fantasy|Mystery	279056317
10	Adventure|Family|Fantasy|Mystery	301956980		Action|Adventure	138869176.3
11	Action|Adventure|Sci-Fi	330249062		Action|Adventure|Western	52483256.75
12	Action|Adventure|Sci-Fi	200069408		Action|Adventure|Family|Fantasy	90106054
13	Action|Adventure	168368427		Action|Adventure|Comedy|Family|Fantasy|Sci-F	88384658.75
14	Action|Adventure|Fantasy	423032628		Adventure|Fantasy	93633631.86
15	Action|Adventure|Western	89289910		Action|Adventure|Drama|History	52646074
16	Action|Adventure|Fantasy|Sci-Fi	291021565		Adventure|Family|Fantasy	133455583.2
17	Action|Adventure|Family|Fantasy	141614023		Action|Adventure|Drama|Romance	154114075.1
18	Action|Adventure|Sci-Fi	623279547		Drama|Romance	32388924.63
19	Action|Adventure|Fantasy	241063875		Action|Adventure|Sci-Fi|Thriller	108151566.4
20	Action|Adventure|Comedy|Family|Fantasy|Sci-Fi	179020854		Action|Adventure|Fantasy|Romance	289279970.3
21	Adventure|Fantasy	255108370		Action|Adventure|Fantasy|Sci-Fi|Thriller	95709585.08
22	Action|Adventure|Fantasy	262030663		Adventure|Animation|Comedy|Family|Fantasy	143316248.9



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.

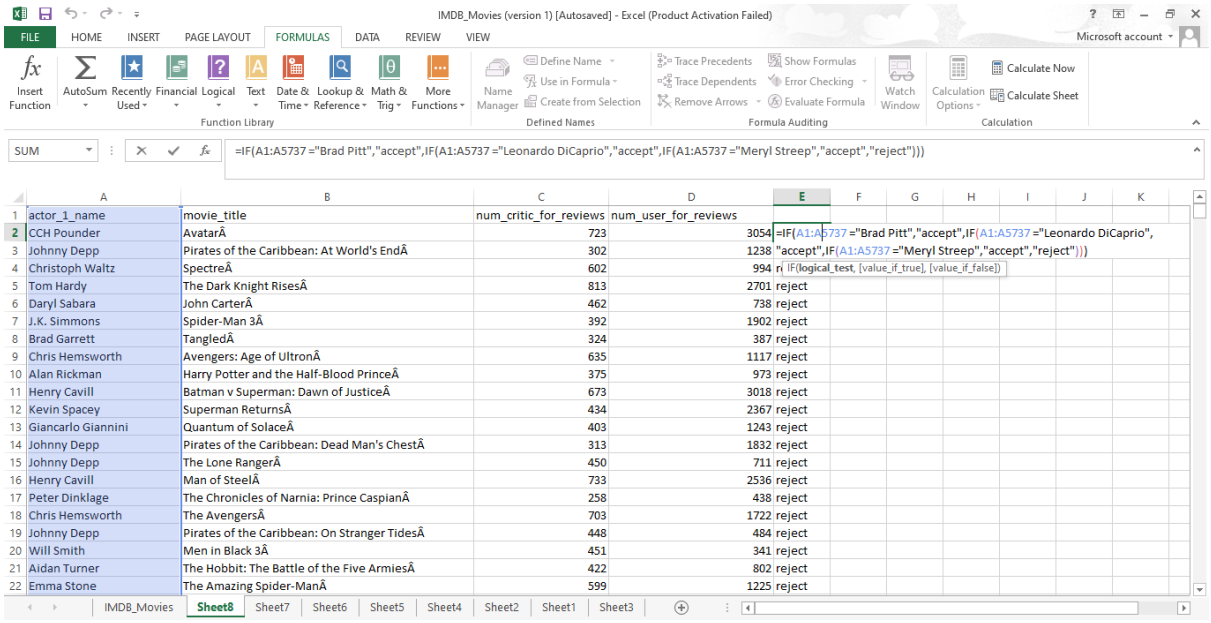
   Append the rows of all these columns and store them in a new column named Combined.

   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.

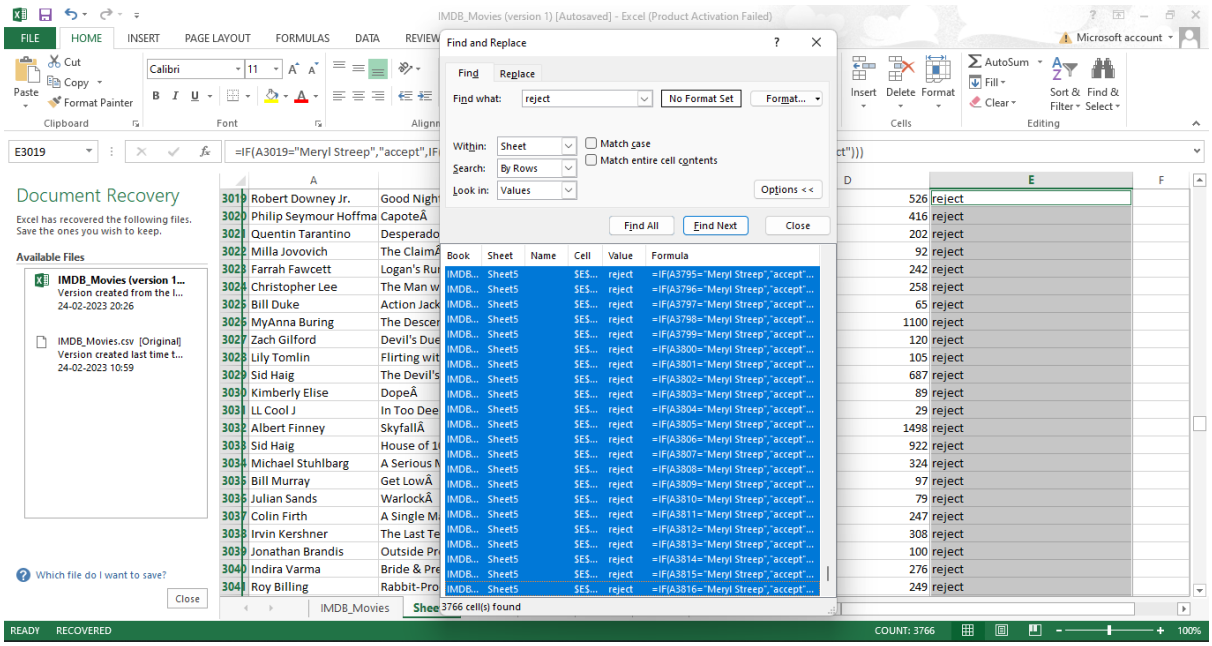
use nested if function to get the data of 3 actors separately

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

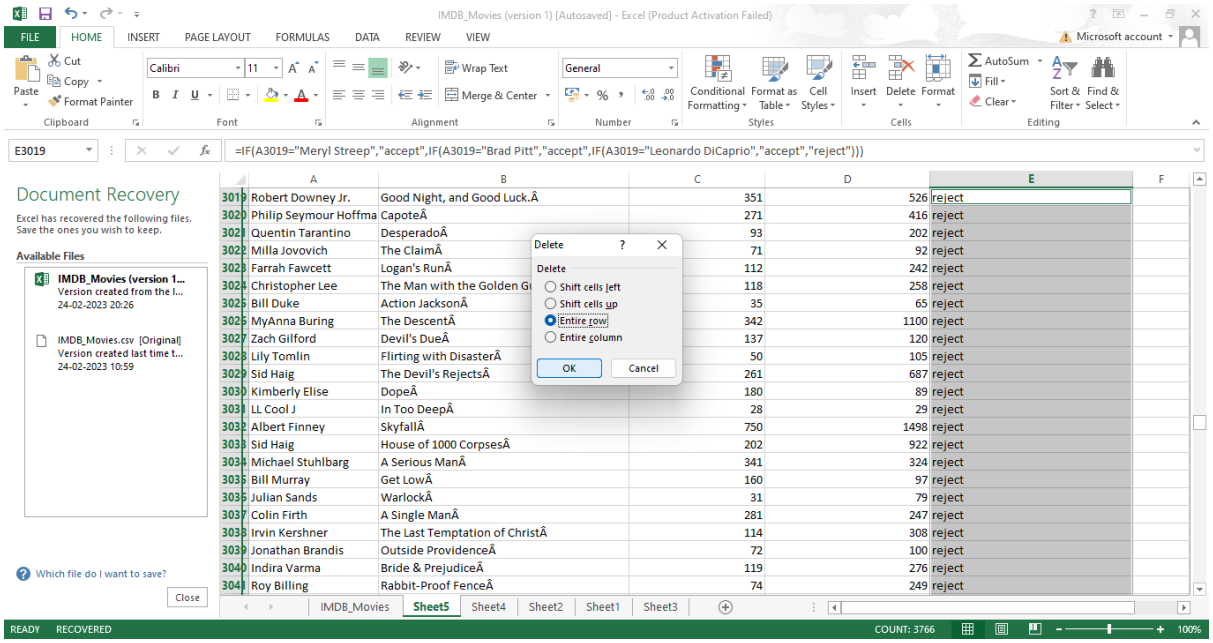


	A	B	C	D	E
1	actor_1_name	movie_title	num_critic_for_reviews	num_user_for_reviews	
2	CCH Pounder	AvatarÂ	723	3054	=IF(A1:A5737 ="Brad Pitt","accept",IF(A1:A5737 ="Leonardo DiCaprio",
3	Johnny Depp	Pirates of the Caribbean: At World's EndÂ	302	1238	"accept",IF(A1:A5737 ="Meryl Streep","accept","reject")))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

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



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



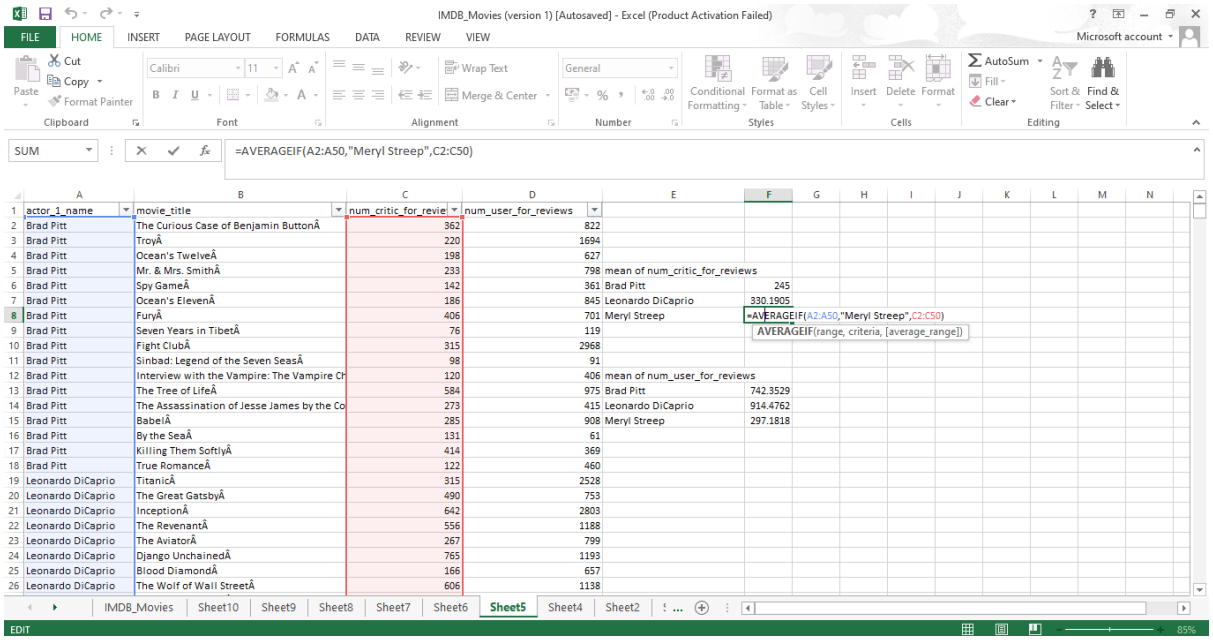
To find the mean of the 3 actors use average if condition

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)

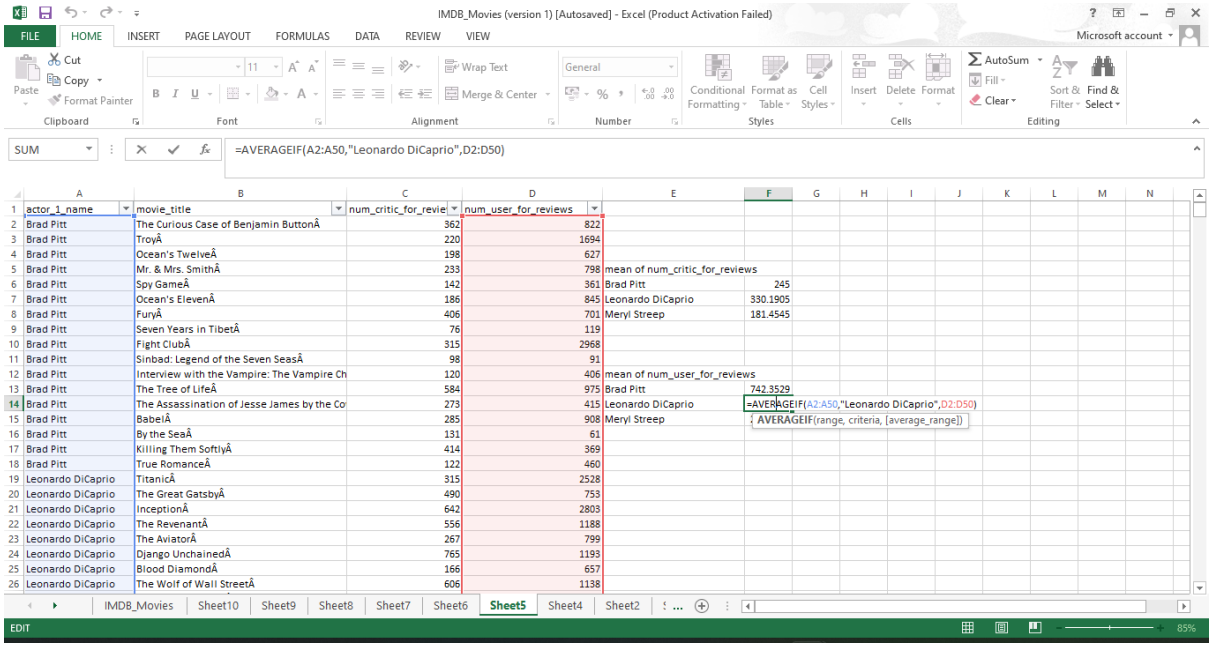


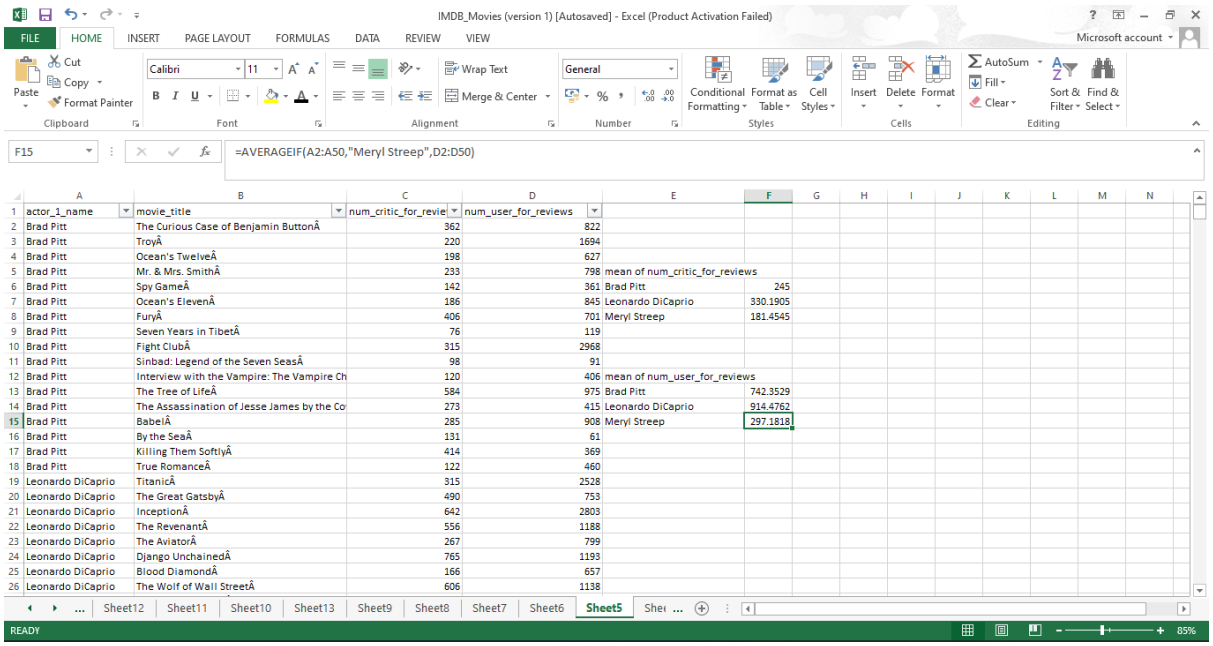
mean of num_user_for_reviews

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

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

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

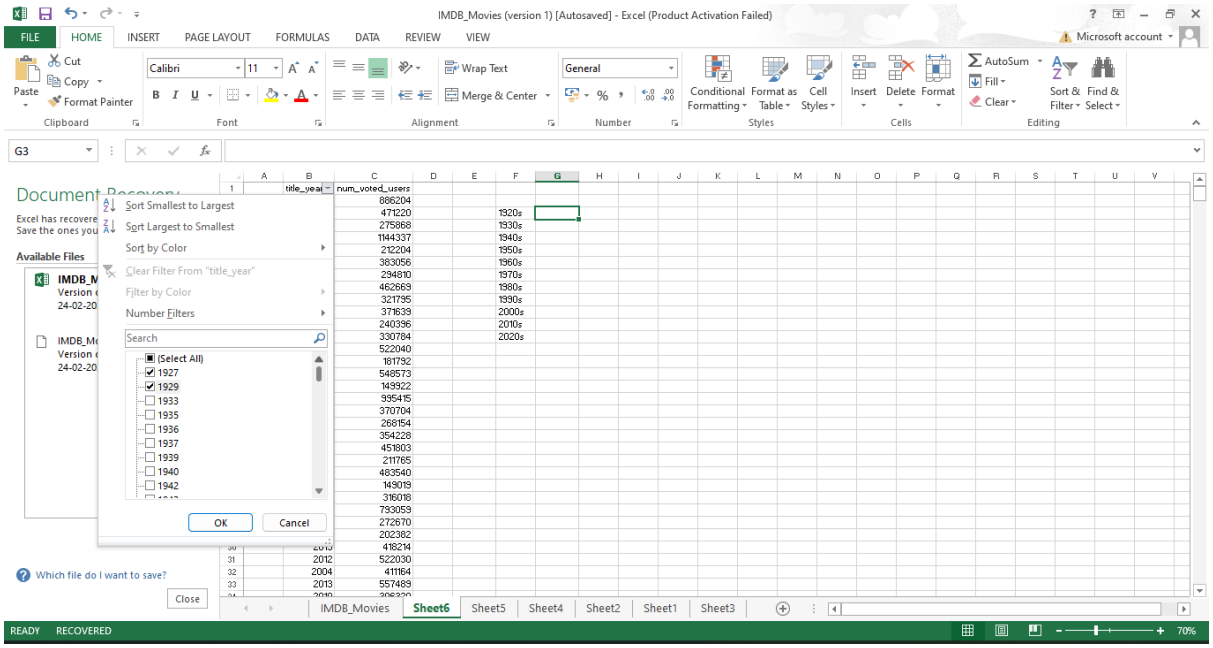


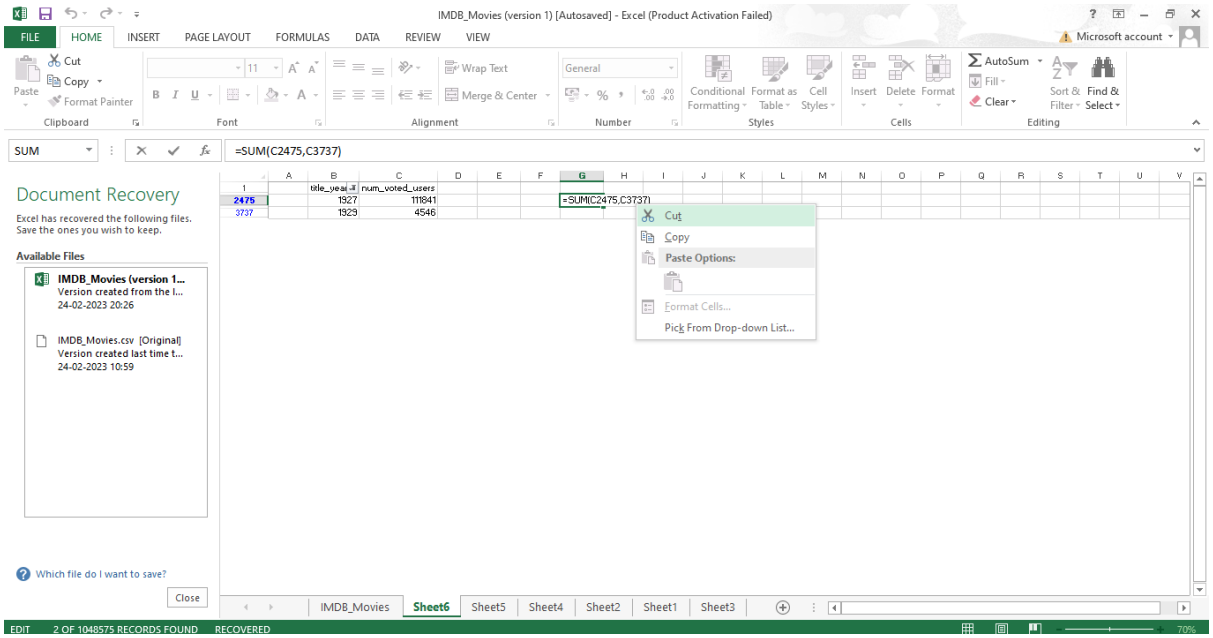


the actors which have the highest mean is Leonardo DiCaprio

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 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

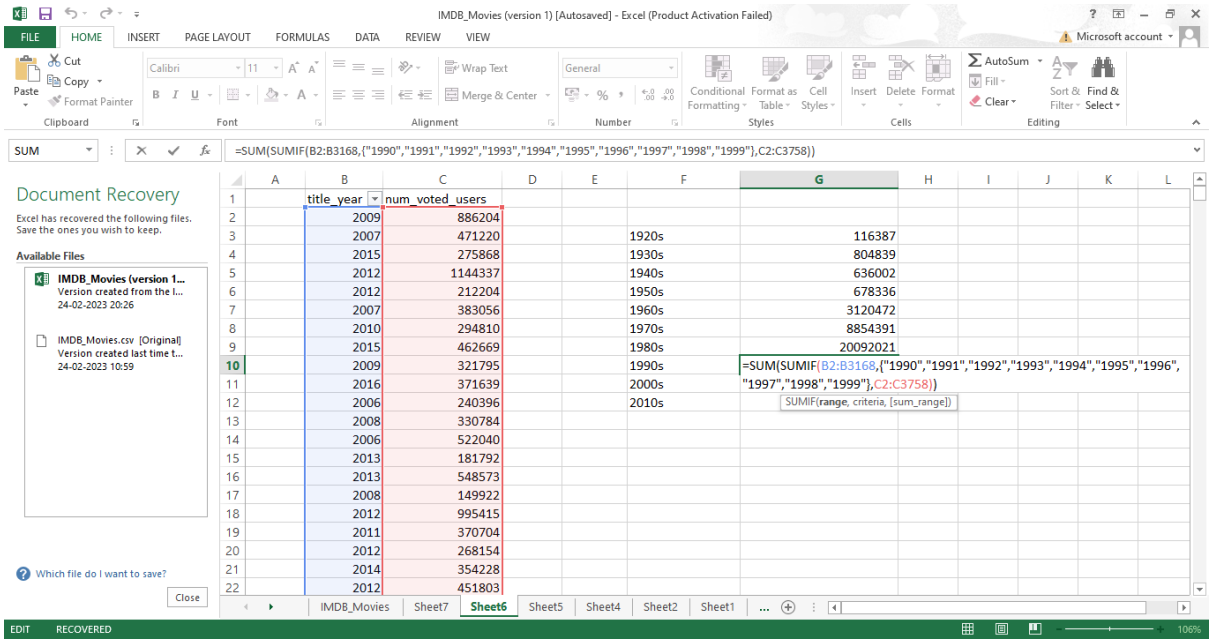




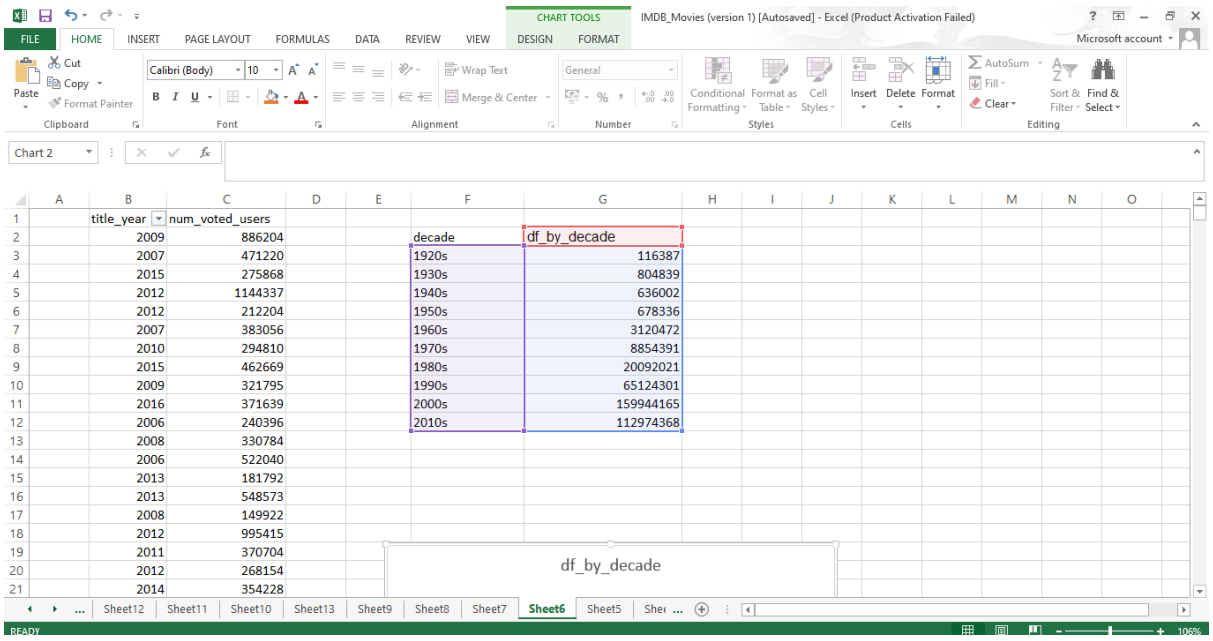
If there are larged number of values to perform the sum function use the sum if function

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

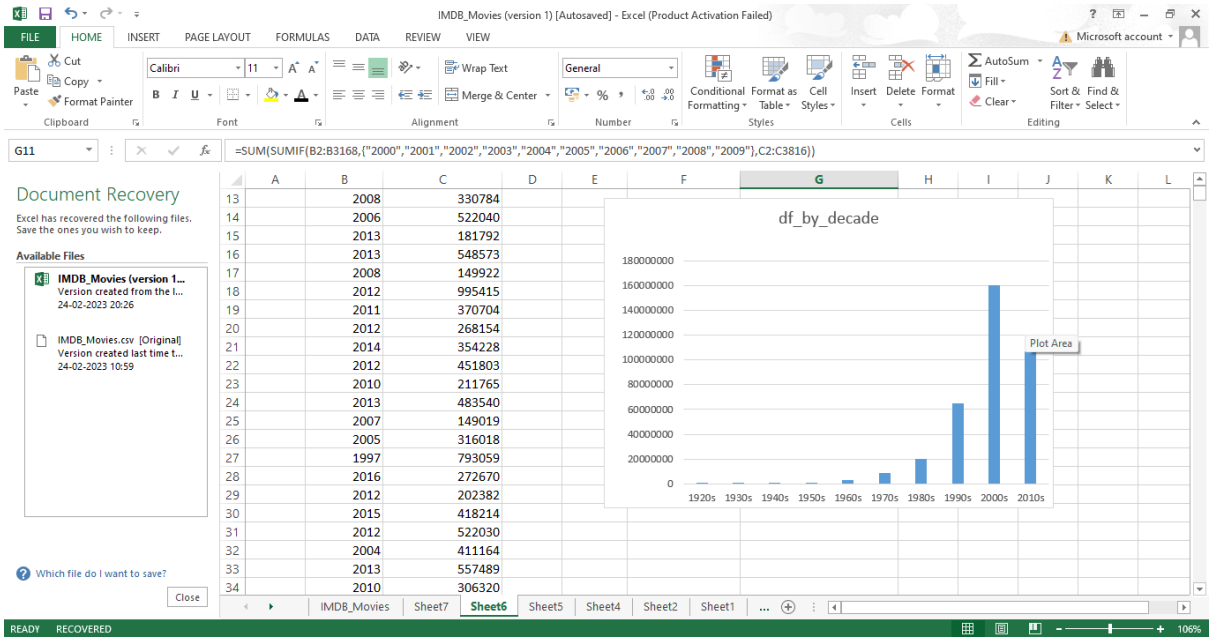
To get the sum of the certain values



	A	B	C	D	E	F	G
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",
11		2016	371639			2000s	"1997","1998","1999"},C2:C3758))
12		2006	240396			2010s	
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				



	A	B	C	D	E	F	G
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				



	A	B	C	D	E	F	G	H	I	J	K	L
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									
23		2010	211765									
24		2013	483540									
25		2007	149019									
26		2005	316018									
27		1997	793059									
28		2016	272670									
29		2012	202382									
30		2015	418214									
31		2012	522030									
32		2004	411164									
33		2013	557489									
34		2010	306320									