

# IMDB Movie Analysis

## Project Description:

This is detailed overview of complete analysis of imdb\_movies. The dataset which used is imdb\_movie dataset. I was required to create a problem which I want to shed some light on.

## Tech Used:

- **Microsoft excel** : excel is used for complete analysis done in this project.
- **google drive**: drive is used for storing this file in it.
- **canva**: canva is used to make this report.

## Approach:

In this project I followed 5 Whys technique to determine its root cause by repeatedly asking the question “Why”. It's also called root cause analysis. which is developed by sakichi toyoda founder of toyota industires. This approach is basically asking why in every step.

## Insights and reports:

I was required to provide a detailed report for the below data record mentioning the answers of the questions that follows:

**a. Cleaning the data:** 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

### Insights:

- In this step I removed all the blank cells.
- there are lots of dropping columns present in this dataset which has no use at all, so I removed all the dropping columns.
- I even removed the duplicate value and made the data set look more appealing.

### Before cleaning:

color	director	num_critc	duration	director_fa	actor_3	facactor_2	naia	actor_1	facgross	genres	actor_1_nai	movie_title	num_voted	cast_total	actor_3_nai	face	numbe	plot	keyw	movie	imd	num_user	reviews	language
Color	James Cam	723	178	0	855 Joel David I	1000	760505847	Action	Adv CCH Pound	Avatar	886204	4834 Wes Studi	0	avata	futu	http://www						3054	English	
Color	Gore Verbi	302	169	563	1000 Orlando Blc	40000	309404152	Action	Adv Johnny De	Pirates of ti	471220	48350 Jack Daven	0	goddess	mh	http://www						1238	English	
Color	Sam Mend	602	148	0	161 Rory Kinne	11000	200074175	Action	Adv Christoph	V/Spectre	275868	11700 Stephanie	1	bomb	espi	http://www						994	English	
Color	Christophe	813	164	22000	23000 Christian Bi	27000	448130642	Action	ThriTom Hardy	The Dark Ki	1144337	106759 Joseph Gor	0	deception	http://www							2701	English	
Color	Doug Walker			131	Rob Walker	131		Documents	Doug Walk	Star Wars: f	8	143	0		http://www									
Color	Andrew Sta	462	132	475	530 Samantha H	640	73058679	Action	Adv Daryl Sabar	John Carter	212204	1873 Polly Walk	1	alien	amer	http://www							738	English
Color	Sam Raimi	392	156	0	4000 James Fran	24000	336530303	Action	Adv J.K. Simmo	Spider-Mar	383056	46055 Kirsten Dur	0	sandman	s	http://www							1902	English
Color	Nathan Gre	324	100	15	284 Donna Mur	799	200807262	Adventure	Brad Garret	Tangled	294810	2036 M.C. Gaine	1	17th centu	http://www								387	English
Color	Joss Whed	635	141	0	19000 Robert Dow	26000	458991599	Action	Adv Chris Hems	Avengers: A	462669	92000 Scarlett Joh	4	artificial in	http://www								1117	English
Color	David Yates	375	153	282	10000 Daniel Rad	25000	301956980	Adventure	Alan Rickm	Harry Potte	321795	58753 Rupert Grin	3	blood	bool	http://www							973	English
Color	Zack Snyder	673	183	0	2000 Lauren Chu	15000	330249062	Action	Adv Henry Cavil	Batman v Si	371639	24450 Alan D. Pur	0	based on cc	http://www								3018	English
Color	Bryan Singe	434	169	0	903 Marlon Brai	18000	200069408	Action	Adv Kevin Spa	Superman f	240396	29991 Frank Lang	0	crystal	epi	http://www							2367	English
Color	Marc Forste	403	106	395	393 Mathieu Ar	451	168368427	Action	Adv Giancarlo	Quantum o	330784	2023 Rory Kinne	1	action hero	http://www								1243	English
Color	Gore Verbi	313	151	563	1000 Oriando Blc	40000	423032628	Action	Adv Johnny De	Pirates of ti	522040	48486 Jack Daven	2	box office	http://www								1832	English
Color	Gore Verbi	450	150	563	1000 Ruth Wilsoi	40000	89289910	Action	Adv Johnny De	The Lone Ri	181792	45757 Tom Wilkin	1	horse	out	http://www							711	English
Color	Zack Snyder	733	143	0	748 Christophe	15000	291021565	Action	Adv Henry Cavil	Man of Ste	548573	20495 Harry Lenni	0	based on cc	http://www								2536	English
Color	Andrew Ad	258	150	80	201 Pierfrances	22000	141614023	Action	Adv Peter Dinkl	The Chroni	149922	22697 Damián Alc	4	brother br	http://www								438	English
Color	Joss Whed	703	173	0	19000 Robert Dow	26000	623279547	Action	Adv Chris Hems	The Avengi	995415	87697 Scarlett Joh	3	alien invasi	http://www								1722	English
Color	Rob Marshi	448	136	252	1000 Sam Claflin	40000	241063875	Action	Adv Johnny De	Pirates of ti	370704	54083 Stephen Gr	4	blackbeard	http://www								484	English
Color	Barry Sonn	451	106	188	718 Michael Stu	10000	179020854	Action	Adv Will Smith	Men in Blak	268154	12572 Nicole Sche	1	alien	crimi	http://www							341	English
Color	Peter Jacks	422	164	0	773 Adam Brow	5000	255108370	Adventure	Aidan Turn	The Hobbit	354228	9152 James Nest	0	army	elf	http://www							802	English
Color	Marc Webb	599	153	464	963 Andrew Gai	15000	262036663	Action	Adv Emma Sto	The Hobbit	451803	28489 Chris Zylka	0	lizard	out	http://www							1225	English
Color	Ridley Scot	343	156	0	738 William Hu	891	105219735	Action	Adv Mark Addy	Robin Hooc	211765	3244 Scott Grime	0	1190s	arch	http://www							546	English
Color	Peter Jacks	509	186	0	773 Adam Brow	5000	258355354	Adventure	Aidan Turn	The Hobbit	483540	9152 James Nest	0	based on cc	http://www								951	English
Color	Chris Weitz	251	113	129	1000 Eva Green	16000	70083519	Adventure	Christophe	The Golden	149019	24106 Kristin Scot	2	children	elf	http://www							666	English
Color	Peter Jacks	446	201	0	84 Thomas Kre	6000	218051260	Action	Adv Naomi Wat	King Kong	316018	7123 Evan Parke	0	animal nar	http://www								2618	English
Color	James Cam	315	194	0	794 Kate Winsl	29000	658672302	Drama	Ron Leonardo	D/Titanic	793059	45223 Gloria Stuar	0	artist	love	http://www							2528	English
Color	Anthony Ri	516	147	94	11000 Scarlett Joh	21000	407197282	Action	Adv Robert Dow	Captain Am	272670	64798 Chris Evans	0	based on cc	http://www								1022	English
Color	Peter Berg	377	131	532	627 Alexander	14000	65173160	Action	Adv Liam Nees	Battleship	202382	26679 Tadanobu	0	box office	http://www								751	English
Color	Colin Teena	684	134	306	9000 Isiah Guti	3000	652177737	Action	Adv Danya Dalia	Superhero M	4919714	8460 Pooja Ch	0	action nar	http://www								1300	English

### After Cleaning:

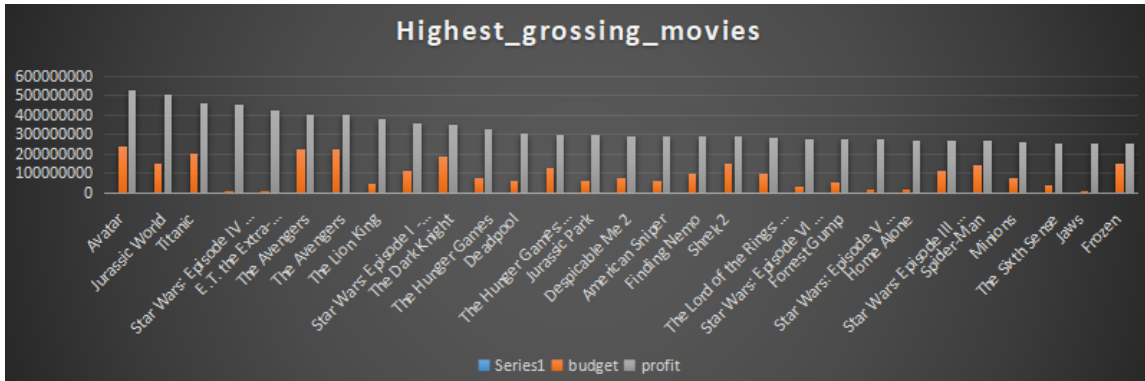
	A	B	C	D	E	F	G	H	
	director_name	num_critic_for_reviv	gross	genres	actor_1_name	movie_title	num_voted_user	num_user_for_review	lang
1	James Cameron	723	760505847	Action Adventure Fantasy Sci-Fi	CGH Pounder	Avatar	886204	3054	Engl
1	Gore Verbinski	302	389404152	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End	471220	1238	Engl
1	Sam Mendes	602	200074175	Action Adventure Thriller	Christoph Waltz	Spectre	275868	994	Engl
1	Christopher Nolan	813	448130642	Action Thriller	Tom Hardy	The Dark Knight Rises	1144337	2701	Engl
1	Andrew Stanton	462	73058679	Action Adventure Sci-Fi	Daryl Sabara	John Carter	212204	738	Engl
1	Sam Raimi	392	336530303	Action Adventure Romance	J.K. Simmons	Spider-Man 3	383056	1902	Engl
1	Nathan Greno	324	200807262	Adventure Animation Comedy Family Fantasy Musical Romance	Brad Garrett	Tangled	294810	387	Engl
1	Joss Whedon	635	458991599	Action Adventure Sci-Fi	Chris Hemsworth	Avengers: Age of Ultron	462669	1117	Engl
0	David Yates	375	301956980	Adventure Family Fantasy Mystery	Alan Rickman	Harry Potter and the Half-Blood Prince	321795	973	Engl
1	Zack Snyder	673	330249062	Action Adventure Sci-Fi	Henry Cavill	Batman v Superman: Dawn of Justice	371639	3018	Engl
2	Bryan Singer	434	200069408	Action Adventure Sci-Fi	Kevin Spacey	Superman Returns	240396	2367	Engl
3	Marc Forster	403	168368427	Action Adventure	Giancarlo Giannini	Quantum of Solace	330784	1243	Engl
4	Gore Verbinski	313	423032628	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: Dead Man's Chest	522040	1832	Engl
5	Gore Verbinski	450	89289910	Action Adventure Western	Johnny Depp	The Lone Ranger	181792	711	Engl
6	Zack Snyder	733	291021565	Action Adventure Fantasy Sci-Fi	Henry Cavill	Man of Steel	548573	2536	Engl
7	Andrew Adamson	258	141614023	Action Adventure Family Fantasy	Peter Dinklage	The Chronicles of Narnia: Prince Caspian	149922	438	Engl
8	Joss Whedon	703	623279547	Action Adventure Sci-Fi	Chris Hemsworth	The Avengers	995415	1722	Engl
9	Rob Marshall	448	241063875	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: On Stranger Tides	370704	484	Engl
0	Barry Sonnenfeld	451	179020854	Action Adventure Comedy Family Fantasy Sci-Fi	Will Smith	Men in Black 3	268154	341	Engl
1	Peter Jackson	422	255108370	Adventure Fantasy	Aidan Turner	The Hobbit: The Battle of the Five Armies	354228	802	Engl
2	Marc Webb	599	262030663	Action Adventure Fantasy	Emma Stone	The Amazing Spider-Man	451803	1225	Engl
3	Ridley Scott	343	105219735	Action Adventure Drama History	Mark Addy	Robin Hood	211765	546	Engl
4	Peter Jackson	509	258355354	Adventure Fantasy	Aidan Turner	The Hobbit: The Desolation of Smaug	483540	951	Engl
5	Chris Weitz	251	70083519	Adventure Family Fantasy	Christopher Lee	The Golden Compass	149019	666	Engl
6	Peter Jackson	446	218051260	Action Adventure Drama Romance	Naomi Watts	King Kong	316018	2618	Engl
7	James Cameron	315	658672302	Drama Romance	Leonardo DiCaprio	Titanic	793059	2528	Engl
8	Anthony Russo	516	407197282	Action Adventure Sci-Fi	Robert Downey Jr.	Captain America: Civil War	272670	1022	Engl
9	Peter Berg	377	65173160	Action Adventure Sci-Fi Thriller	Liam Neeson	Battleship	202382	751	Engl
0	Colin Trevorrow	644	652177271	Action Adventure Sci-Fi Thriller	Bryce Dallas Howard	Jurassic World	418214	1290	Engl
1	Sam Mendes	764	304260777	Action Adventure Thriller	Albert Finney	Clouds	679020	1468	Engl

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

**Your task:** Find the movies with the highest profit?

**Insights:**

- So basically profit is all about gross - budget. In this step I simply reduce the gross from budget and kept them in separate column.
- on the analysis of this movies, we came to know about the top highest grossed movies ever.
- we can able to see the difference between the budget and profit for these very clearly with the help of charts.



**c. Top 250:** Create a new column IMDb\_Top\_250 and store the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb\_score). Also make sure that for all of these movies, the num\_voted\_users is greater than 25,000. Also add a Rank column containing the values 1 to 250 indicating the ranks of the corresponding films.

Extract all the movies in the IMDb\_Top\_250 column which are not in the English language and store them in a new column named Top\_Foreign\_Lang\_Film. You can use your own imagination also!

**Your task:** Find IMDB Top 250

**Insights:**

- In this task we came to know about the imdb top 250 movies. about that movies we only group movies only in English language and rest of them in other column.
- These top movies have filtered based on the num\_voted\_users. the movies which has num\_voted\_users is more than 25000 had made into this list.

Top\_imdb\_movies:

movie_title	imdb_score	num_voted_users	rank	T:
The Shawshank Redemption	9.3	886204	1	
The Godfather	9.2	471220	2	
The Dark Knight	9	275868	3	
The Godfather: Part II	9	1144337	4	
The Lord of the Rings: The Return of the King	8.9	212204	5	
Schindler's List	8.9	383056	6	
Pulp Fiction	8.9	294810	7	
The Good, the Bad and the Ugly	8.9	462669	8	
Inception	8.8	321795	9	
The Lord of the Rings: The Fellowship of the Ring	8.8	371639	10	
Fight Club	8.8	240396	11	
Forrest Gump	8.8	330784	12	
Star Wars: Episode V - The Empire Strikes Back	8.8	522040	13	
The Lord of the Rings: The Two Towers	8.7	181792	14	
The Matrix	8.7	548573	15	
Goodfellas	8.7	149922	16	
Star Wars: Episode IV - A New Hope	8.7	995415	17	
One Flew Over the Cuckoo's Nest	8.7	370704	18	
City of God	8.7	268154	19	
Seven Samurai	8.7	354228	20	
Interstellar	8.6	451803	21	
Saving Private Ryan	8.6	211765	22	
Se7en	8.6	483540	23	
The Silence of the Lambs	8.6	149019	24	
Spirited Away	8.6	316018	25	
American History X	8.6	793059	26	
The Usual Suspects	8.6	272670	27	
Modern Times	8.6	202382	28	
The Dark Knight Rises	8.5	418214	29	
Gladiator	8.5	572020	30	

Top\_imdb\_movies in foreign languages:

movie_title	imdb_score	num_voted_users	language	T:	rank	T:
The Good, the Bad and the Ugly	8.9	503509	Italian		8	
City of God	8.7	533200	Portuguese		19	
Seven Samurai	8.7	229012	Japanese		20	
Spirited Away	8.6	417971	Japanese		25	
The Lives of Others	8.5	259379	German		45	
Children of Heaven	8.5	27882	Persian		46	
Amélie	8.4	534262	French		49	
Princess Mononoke	8.4	221552	Japanese		52	
Das Boot	8.4	168203	German		56	
Oldboy	8.4	356181	Korean		58	
A Separation	8.4	151812	Persian		60	
Metropolis	8.3	111841	German		74	
Downfall	8.3	248354	German		76	
The Hunt	8.3	170155	Danish		83	
Howl's Moving Castle	8.2	214091	Japanese		99	
Pan's Labyrinth	8.2	467234	Spanish		101	
Incendies	8.2	80429	French		103	
The Secret in Their Eyes	8.2	131831	Spanish		105	
The Act of Killing	8.2	23836	Indonesian		109	
The Sea Inside	8.1	64556	Spanish		141	
Tae Guk Gi: The Brotherhood of War	8.1	31943	Korean		142	
Akira	8.1	106160	Japanese		144	
Elite Squad	8.1	81644	Portuguese		145	
Amores Perros	8.1	173551	Spanish		154	
The Celebration	8.1	65951	Danish		156	
My Name Is Khan	8	69759	Hindi		202	

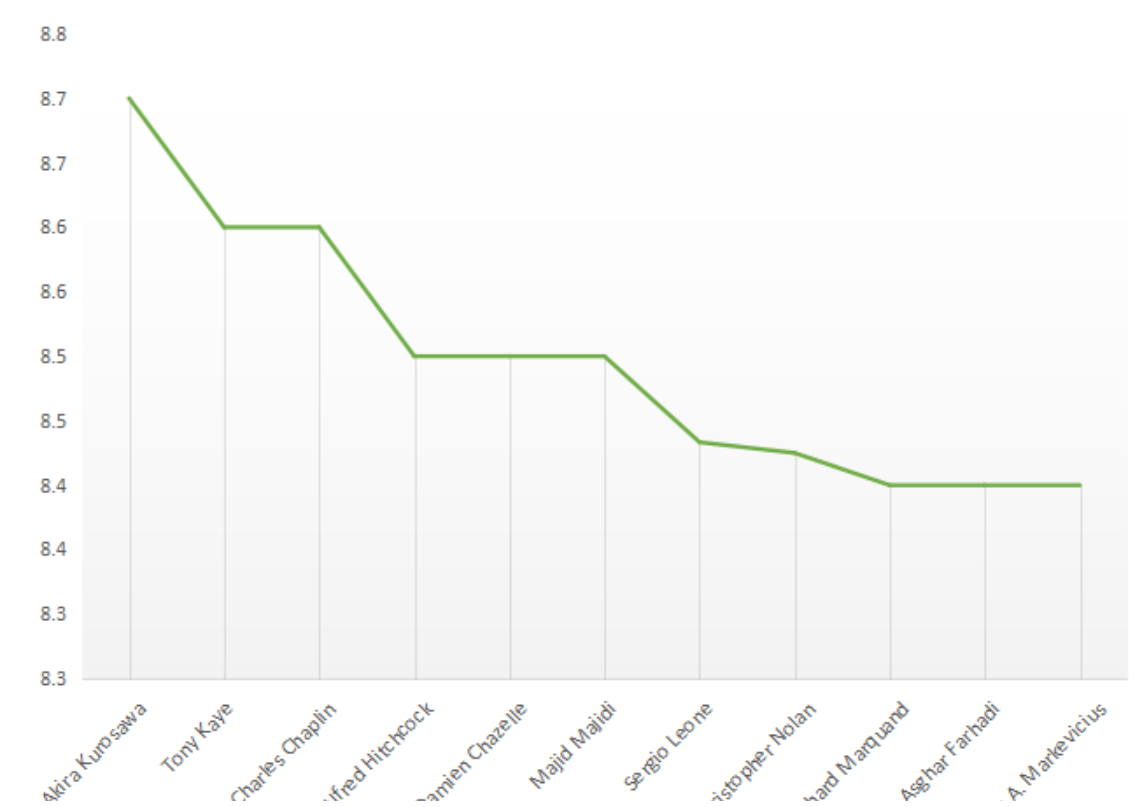
Persepolis	8	70194	French	204
Central Station	8	28951	Portuguese	210
Waltz with Bashir	8	46107	Hebrew	213

**d. Best Directors:** Group 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 top 10 director.

**Your task:** Find the best directors

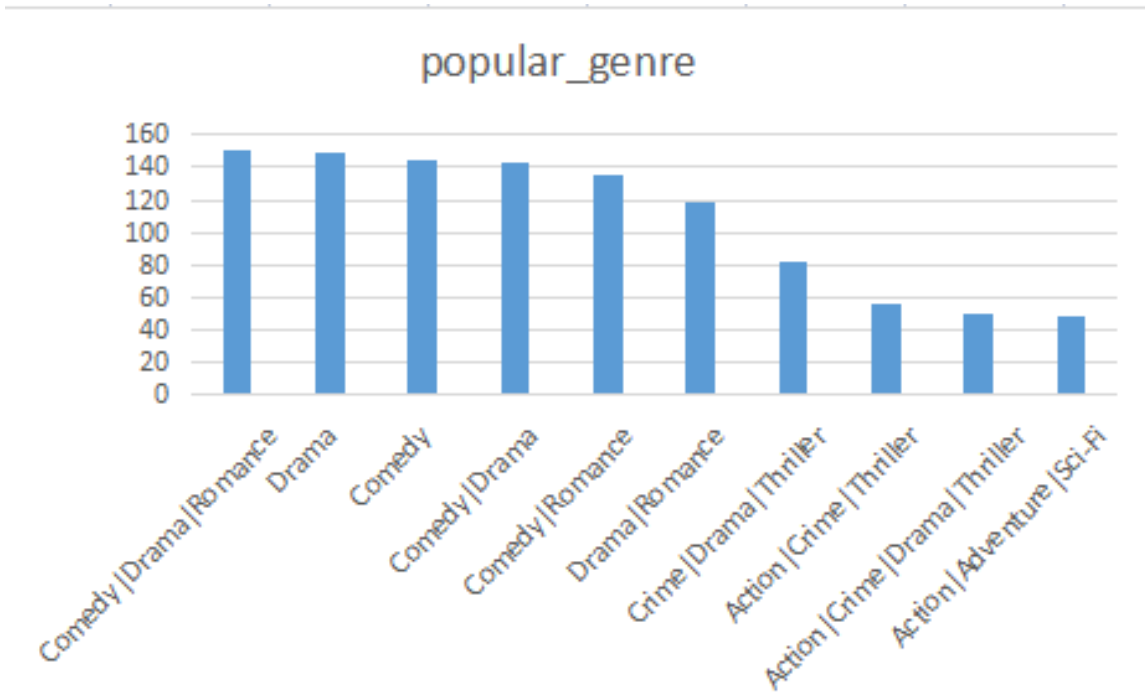
director_name	imdb_score
Akira Kurosawa	8.7
Tony Kaye	8.6
Charles Chaplin	8.6
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Sergio Leone	8.4
Christopher Nolan	8.4
Richard Marquand	8.4
Asghar Farhadi	8.4
Marius A. Markevicius	8.4

top\_10\_director



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

E		F	
genres		count	
Comedy   Drama   Romance		150	
Drama		149	
Comedy		145	
Comedy   Drama		143	
Comedy   Romance		135	
Drama   Romance		119	
Crime   Drama   Thriller		82	
Action   Crime   Thriller		56	
Action   Crime   Drama   Thriller		50	
Action   Adventure   Sci-Fi		48	



**F: Charts:** Create three new columns namely, Meryl\_Streep, Leo\_Caprio, and Brad\_Pitt which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors. Use only the actor\_1\_name column for extraction. Also, make sure that you

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

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

Group the combined column using the actor\_1\_name column.

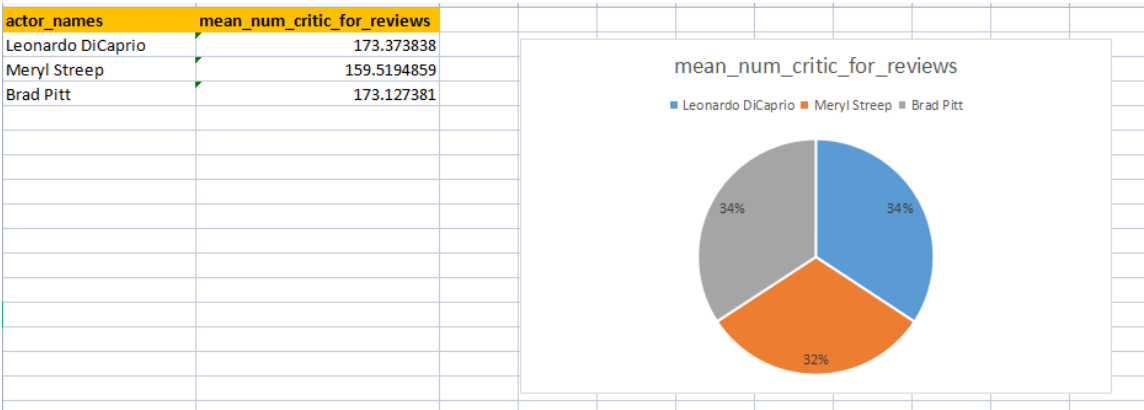
Find the mean of the num\_critic\_for\_reviews and num\_users\_for\_review and identify the actors which have the highest mean.

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

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

I	J	K
Leonardo DiCaprio	Meryl Streep	Brad Pitt
Titanic	It's Complicated	The Curious Case of Benjamin Button
The Great Gatsby	The River Wild	Troy
Inception	Julie & Julia	Ocean's Twelve
The Revenant	The Devil Wears Prada	Mr. & Mrs. Smith
The Aviator	Lions for Lambs	Spy Game
Django Unchained	Out of Africa	Ocean's Eleven
Blood Diamond	Hope Springs	Fury
The Wolf of Wall Street	One True Thing	Seven Years in Tibet
Gangs of New York	The Hours	Fight Club
The Departed	The Iron Lady	Sinbad: Legend of the Seven Seas
Shutter Island	A Prairie Home Companion	Interview with the Vampire: The Vampire Chronicles
Body of Lies		The Tree of Life
Catch Me If You Can		The Assassination of Jesse James by the Coward Robert Fo
The Beach		Babel
Revolutionary Road		By the Sea
The Man in the Iron Mask		Killing Them Softly
J. Edgar		True Romance
The Quick and the Dead		
Marvin's Room		
Romeo + Juliet		
The Great Gatsby		

critic favorite actors:



Audience favorite actors:





