

IMDB Movie Analysis

Final Project-1

Description:

For your Final Project, we are providing you with dataset having various columns of different IMDB Movies. You are required to Frame the problem. For this task, you will need to define a problem you want to shed some light on.

We can do this by asking 'What?' This is where you frame the problem i.e. What is the problem?

Use these questions to guide your thinking:

- What do you see happening?
- What is your hypothesis for the cause of the problem? (This will be broadly based on intuition initially)
- What is the impact of the problem on stakeholders?
- What is the impact of the problem not being solved?

Answering these questions will help you define a problem you are trying to solve and will allow you to find the right data to solve it.

Approach: After carefully understanding the data and problem statement, I went through each problem statement one by one and using tables, pie charts and pivot tables I solved each problem statement.

Tech-Stack Used: Microsoft Excel

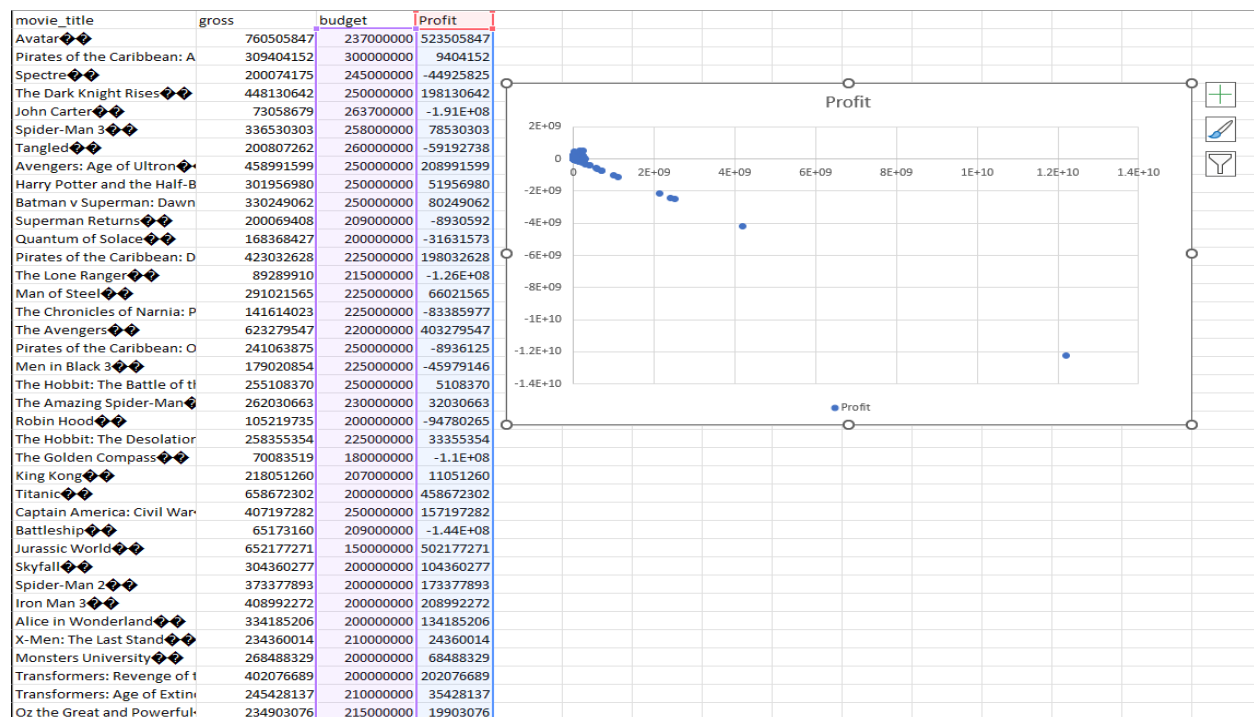
Insights: After completing this project I understood what it means to clean the data and analyze which information should be used for analyzing.

Result:

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.)

	director_name	num_critc_duration	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_for_revie	language	country	budget	title_year	imdb_score	
1	James Cameron	723	178	760505847	Action Adventure	CCH Pounder	Avatar	886204	3054	English	USA	237000000	2009	7.9
2	Gore Verbinski	302	169	309404152	Action Adventure	Johnny Depp	Pirates of the Caribbean: A	471220	1238	English	USA	300000000	2007	7.1
3	Sam Mendes	602	148	200074175	Action Adventure	Christoph Waltz	Spectre	275868	994	English	UK	245000000	2015	6.8
4	Christopher Nol	813	164	448130642	Action Thriller	Tom Hardy	The Dark Knight Rises	1144337	2701	English	USA	250000000	2012	8.5
5	Andrew Stanton	462	132	73058679	Action Adventure	Daryl Sabara	John Carter	212204	738	English	USA	263700000	2012	6.6
6	Sam Raimi	392	156	336530303	Action Adventure	J.K. Simmons	Spider-Man 3	383056	1902	English	USA	258000000	2007	6.2
7	Nathan Greno	324	100	200807262	Adventure Animati	Brad Garrett	Tangled	294810	387	English	USA	260000000	2010	7.8
8	Joss Whedon	635	141	458991599	Action Adventure	Chris Hemsworth	Avengers: Age of Ultron	462669	1117	English	USA	250000000	2015	7.5
9	David Yates	375	153	301956980	Adventure Family	Alan Rickman	Harry Potter and the Half-B	321795	973	English	UK	250000000	2009	7.5
10	Zack Snyder	673	183	330249062	Action Adventure	Henry Cavill	Batman v Superman: Dawn	371639	3018	English	USA	250000000	2016	6.9
11	Bryan Singer	434	169	200069408	Action Adventure	Kevin Spacey	Superman Returns	240396	2367	English	USA	209000000	2006	6.1
12	Marc Forster	403	106	168368427	Action Adventure	Giancarlo Giannini	Quantum of Solace	330784	1243	English	UK	200000000	2008	6.7
13	Gore Verbinski	313	151	423032628	Action Adventure	Johnny Depp	Pirates of the Caribbean: D	522040	1832	English	USA	225000000	2006	7.3
14	Gore Verbinski	450	150	89289910	Action Adventure	Johnny Depp	The Lone Ranger	181792	711	English	USA	215000000	2013	6.5
15	Zack Snyder	733	143	291021565	Action Adventure	Henry Cavill	Man of Steel	548573	2536	English	USA	225000000	2013	7.2
16	Andrew Adamsc	258	150	141614023	Action Adventure	Peter Dinklage	The Chronicles of Narnia: P	149922	438	English	USA	225000000	2008	6.6
17	Joss Whedon	703	173	623279547	Action Adventure	Chris Hemsworth	The Avengers	995415	1722	English	USA	220000000	2012	8.1
18	Rob Marshall	448	136	241063875	Action Adventure	Johnny Depp	Pirates of the Caribbean: O	370704	484	English	USA	250000000	2011	6.7
19	Barry Sonnenfel	451	106	179020854	Action Adventure	Will Smith	Men in Black 3	268154	941	English	USA	225000000	2012	6.8
20	Peter Jackson	422	164	255108370	Adventure Fantasy	Aidan Turner	The Hobbit: The Battle of tl	354228	802	English	New Zeal	250000000	2014	7.5
21	Marc Webb	599	153	262030663	Action Adventure	Emma Stone	The Amazing Spider-Man	451803	1225	English	USA	230000000	2012	7
22	Ridley Scott	343	156	105219735	Action Adventure	Mark Addy	Robin Hood	211765	546	English	USA	200000000	2010	6.7
23	Peter Jackson	509	186	258355354	Adventure Fantasy	Aidan Turner	The Hobbit: The Desolator	483540	951	English	USA	225000000	2013	7.9
24	Chris Weitz	251	113	70083519	Adventure Family	Christopher Lee	The Golden Compass	149019	666	English	USA	180000000	2007	6.1
25	Peter Jackson	446	201	218051260	Action Adventure	Naomi Watts	King Kong	316018	2618	English	New Zeal	207000000	2005	7.2
26	James Cameron	315	194	658672302	Drama Romance	Leonardo DiCaprio	Titanic	793059	2528	English	USA	200000000	1997	7.7
27	Anthony Russo	516	147	407197282	Action Adventure	Robert Downey Jr.	Captain America: Civil War	272670	1022	English	USA	250000000	2016	8.2
28	Peter Berg	377	131	65173160	Action Adventure	Liam Neeson	Battleship	202382	751	English	USA	209000000	2012	5.9
29	Colin Trevorrow	644	124	652177271	Action Adventure	Bryce Dallas Howard	Jurassic World	418214	1290	English	USA	150000000	2015	7
30	Sam Mendes	750	143	304360277	Action Adventure	Albert Finney	Skyfall	522030	1498	English	UK	200000000	2012	7.8
31	Sam Raimi	300	135	373377893	Action Adventure	J.K. Simmons	Spider-Man 2	411164	1303	English	USA	200000000	2004	7.3
32	Shane Black	608	195	408992272	Action Adventure	Robert Downey Jr.	Iron Man 3	557489	1187	English	USA	200000000	2013	7.2
33	Tim Burton	451	108	334185206	Adventure Family	Johnny Depp	Alice in Wonderland	306320	736	English	USA	200000000	2010	6.5
34	Brett Ratner	334	104	234360014	Action Adventure	Hugh Jackman	X-Men: The Last Stand	383427	1912	English	Canada	210000000	2006	6.8
35	Dan Scanlon	376	104	268488329	Adventure Animati	Steve Buscemi	Monsters University	235025	265	English	USA	200000000	2013	7.3
36	Michael Bay	366	150	402076689	Action Adventure	Glenn Morshower	Transformers: Revenge of t	323207	1439	English	USA	200000000	2009	6
37	Michael Bay	378	165	245428137	Action Adventure	Blinging Li	Transformers: Age of Extin	242420	918	English	USA	210000000	2014	5.7
38	Sam Raimi	525	130	234903076	Adventure Family	Tim Holmes	Oz the Great and Powerfu	175409	511	English	USA	215000000	2013	6.4

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.



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.

1	IMDb_Top_250	num_voted_users	language	imdb_score	Rank		
2	The Shawshank Redemption	1689764	English	9.3	1		
3	The Godfather	1155770	English	9.2	2		
4	The Godfather: Part II	790926	English	9	3		
5	The Dark Knight	1676169	English	9	4		
6	The Lord of the Rings: The Fellowship Ring	1215718	English	8.9	5		
7	The Good, the Bad and the Ugly	503509	Italian	8.9	6		
8	Schindler's List	865020	English	8.9	7		
9	Pulp Fiction	1324680	English	8.9	8		
10	The Lord of the Rings: The Two Towers	1238746	English	8.8	9		
11	Star Wars: Episode V - The Empire Strikes Back	837759	English	8.8	10		
12	Inception	1468200	English	8.8	11		
13	Forrest Gump	1251222	English	8.8	12		
14	Fight Club	1347461	English	8.8	13		
15	The Matrix	1217752	English	8.7	14		
16	The Lord of the Rings: The Return of the King	1100446	English	8.7	15		
17	Star Wars: Episode IV - A New Hope	911097	English	8.7	16		
18	Seven Samurai	229012	Japanese	8.7	17		
19	One Flew Over the Cuckoo's Nest	680041	English	8.7	18		
20	Goodfellas	728685	English	8.7	19		
21	City of God	533200	Portuguese	8.7	20		
22	The Usual Suspects	740918	English	8.6	21		
23	The Silence of the Lambs	887467	English	8.6	22		
24	Spirited Away	417971	Japanese	8.6	23		
25	Se7en	1023511	English	8.6	24		
26	Saving Private Ryan	881236	English	8.6	25		
27	Modern Times	143086	English	8.6	26		
28	Interstellar	928227	English	8.6	27		
29	American History X	782437	English	8.6	28		
30	Whiplash	399138	English	8.5	29		
31	The Prestige	844052	English	8.5	30		
32	The Pianist	497946	English	8.5	31		
33	The Lives of Others	259379	German	8.5	32		
34	The Lion King	644348	English	8.5	33		
35	The Green Mile	782610	English	8.5	34		
36	The Departed	873649	English	8.5	35		
37	The Dark Knight Rises	1144337	English	8.5	36		
38	Terminator 2: Judgment Day	744891	English	8.5	37		
39	Samsara	22457	None	8.5	38		

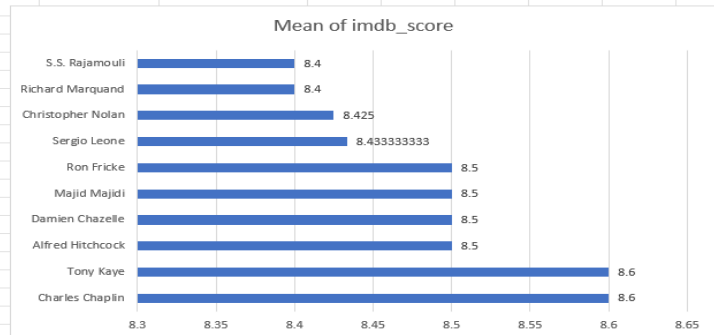
Extract all the movies in the IMDb_Top_250 column which are not in the English language and store them in a new column named Top_Foreign_Lang_Film. You can use your own imagination also!

1	Top_Foreign_Lang_Film	num_voted_use	language	imdb_score	Rank
7	The Good, the Bad and the	503509	Italian	8.9	6
18	Seven Samurai	229012	Japanese	8.7	17
21	City of God	533200	Portuguese	8.7	20
24	Spirited Away	417971	Japanese	8.6	23
33	The Lives of Others	259379	German	8.5	32
39	Samsara	22457	None	8.5	38
45	Children of Heaven	27882	Persian	8.5	44
54	Princess Mononoke	221552	Japanese	8.4	53
56	Oldboy	356181	Korean	8.4	55
58	Das Boot	168203	German	8.4	57
60	Baahubali: The Beginning	62756	Telugu	8.4	59
61	Amélie	534262	French	8.4	60
64	A Separation	151812	Persian	8.4	63
70	The Hunt	170155	Danish	8.3	69
79	Metropolis	111841	German	8.3	78
87	Downfall	248354	German	8.3	86
96	The Secret in Their Eyes	131831	Spanish	8.2	95
99	The Act of Killing	23836	Indonesian	8.2	98
100	Pan's Labyrinth	467234	Spanish	8.2	99
104	Incendies	80429	French	8.2	103
105	Howl's Moving Castle	214091	Japanese	8.2	104
121	The Sea Inside	64556	Spanish	8.1	120
128	The Celebration	65951	Danish	8.1	127
133	Tae Guk Gi: The Brotherhood	31943	Korean	8.1	132
155	Elite Squad	81644	Portuguese	8.1	154
161	Amores Perros	173551	Spanish	8.1	160
162	Akira	106160	Japanese	8.1	161
167	Waltz with Bashir	46107	Hebrew	8	166
186	Persepolis	70194	French	8	185
188	My Name Is Khan	69759	Hindi	8	187
205	Central Station	28951	Portuguese	8	204
218	A Fistful of Dollars	147566	Italian	8	217
220	Veer-Zaara	34449	Hindi	7.9	219
225	The Second Mother	7025	Portuguese	7.9	224
234	The Chorus	44151	French	7.9	233
241	Nine Queens	38215	Spanish	7.9	240
246	Letters from Iwo Jima	132149	Japanese	7.9	245
250	Hero	149414	Mandarin	7.9	249
251	Hero	149414	Mandarin	7.9	250

Best Directors: TGroup the column using the director_name column.

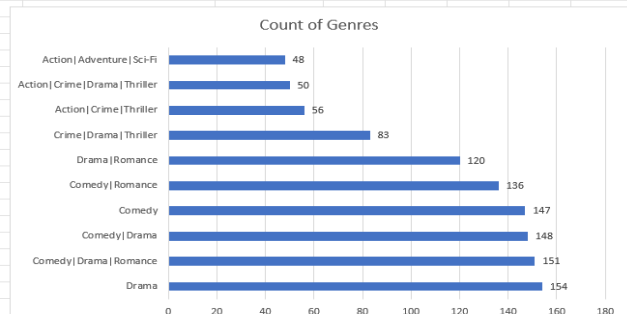
Find out the top 10 directors for whom the mean of imdb_score is the highest and store them in a new column top10director. In case of a tie in IMDb score between two directors, sort them alphabetically.

Row Labels	Average of imdb_score	Top 10 Directors	Mean of imdb_score
Tony Kaye	8.6	Charles Chaplin	8.6
Charles Chaplin	8.6	Tony Kaye	8.6
Ron Fricke	8.5	Alfred Hitchcock	8.5
Majid Majidi	8.5	Damien Chazelle	8.5
Damien Chazelle	8.5	Majid Majidi	8.5
Alfred Hitchcock	8.5	Ron Fricke	8.5
Sergio Leone	8.433333333	Sergio Leone	8.433333333
Christopher Nolan	8.425	Christopher Nolan	8.425
S.S. Rajamouli	8.4	Richard Marquand	8.4
Richard Marquand	8.4	S.S. Rajamouli	8.4
Marius A. Markevicius	8.4		
Asghar Farhadi	8.4		
Lenny Abrahamson	8.3		
Lee Unkrich	8.3		
Fritz Lang	8.3		
Billy Wilder	8.3		
Pete Docter	8.233333333		
Hayao Miyazaki	8.225		
Quentin Tarantino	8.2		
Joshua Oppenheimer	8.2		
Juan José Campanella	8.2		
Elia Kazan	8.2		
George Roy Hill	8.2		
Victor Fleming	8.15		
Milos Forman	8.133333333		
Tim Miller	8.1		
William Wyler	8.1		
Terry George	8.1		
Michael Roemer	8.1		
Michael Wadleigh	8.1		
Je-kyu Kang	8.1		
David Singleton	8.1		
Akira Kurosawa	8.1		
Stephen Chbosky	8		
Vincent Paronnaud	8		
Michel Hazanavicius	8		
David Lean	8		
Ari Folman	8		



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

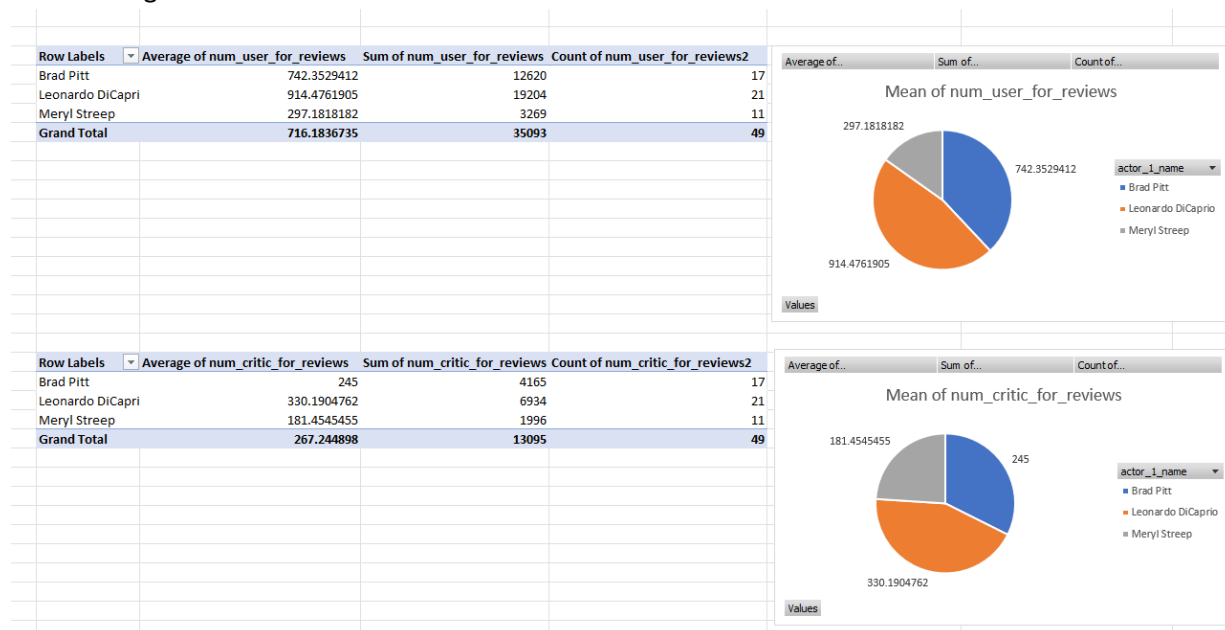
Row Labels	Count of genres	Popular Genres	Count of Genres
Drama	154	Drama	154
Comedy Drama Romance	151	Comedy Drama Romance	151
Comedy Drama	148	Comedy Drama	148
Comedy	147	Comedy	147
Comedy Romance	136	Comedy Romance	136
Drama Romance	120	Drama Romance	120
Crime Drama Thriller	83	Crime Drama Thriller	83
Action Crime Thriller	56	Action Crime Thriller	56
Action Crime Drama Thriller	50	Action Crime Drama Thriller	50
Action Adventure Sci-Fi	48	Action Adventure Sci-Fi	48
Comedy Crime	47		
Action Adventure Thriller	45		
Horror	44		
Drama Thriller	42		
Crime Drama Mystery Thriller	42		
Crime Drama	41		
Horror Thriller	36		
Action Adventure Sci-Fi Thriller	34		
Horror Mystery Thriller	33		
Drama Mystery Thriller	30		
Biography Drama	30		
Adventure Animation Comedy Family Fantasy	28		
Action Comedy Crime	27		
Horror Mystery	25		
Action Adventure Fantasy	25		
Biography Drama Sport	23		
Drama Sport	22		
Drama Music	22		
Action Thriller	22		
Documentary	21		
Adventure Animation Comedy Family	21		
Action Comedy Crime Thriller	21		
Biography Drama History	20		
Comedy Family	20		
Action Crime Drama Mystery Thriller	19		
Action Drama Thriller	18		
Adventure Family Fantasy	18		
Comedy Crime Drama	17		



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.

movie_title	actor_1_name	Meryl_Streep	Leo_Caprio	Brad_Pitt	Combined	Group By
Avatar	CCH Pounder		Titanic	The Curious Case of Benjamin Button	It's Complicated	Meryl_Streep
Pirates of the Caribbean: A Johnny Depp		The River Wild	The Great Gatsby	Troy	The River Wild	Meryl_Streep
Spectre	Christoph Waltz	Julie & Julia	Inception	Ocean's Twelve	Julie & Julia	Meryl_Streep
The Dark Knight Rises	Tom Hardy	The Devil Wears Prada	The Revenant	Mr. & Mrs. Smith	The Devil Wears Prada	Meryl_Streep
John Carter	Daryl Sabara	Lions for Lambs	The Aviator	Spy Game	Lions for Lambs	Meryl_Streep
Spider-Man 3	J.K. Simmons	Out of Africa	Django Unchained	Ocean's Eleven	Out of Africa	Meryl_Streep
Tangled	Brad Garrett	Hope Springs	Blood Diamond	Fury	Hope Springs	Meryl_Streep
Avengers: Age of Ultron	Chris Hemsworth	One True Thing	The Wolf of Wall Street	Seven Years in Tibet	One True Thing	Meryl_Streep
Harry Potter and the Half-Blood Prince	Alan Rickman	The Hours	Gangs of New York	Fight Club	The Hours	Meryl_Streep
Batman v Superman: Dawn of Justice	Henry Cavill	The Iron Lady	The Departed	Sinbad: Legend of the Seven Seas	The Iron Lady	Meryl_Streep
Superman Returns	Kevin Spacey	A Prairie Home Companion	Shutter Island	Interview with the Vampire	A Prairie Home Companion	Meryl_Streep
Quantum of Solace	Giancarlo Giannini		Body of Lies	The Tree of Life	Titanic	Leo_Caprio
Pirates of the Caribbean: On Stranger Tides	Johnny Depp		Catch Me If You Can	The Assassination of Jesse James by the Coward Jim Bowie	The Great Gatsby	Leo_Caprio
The Lone Ranger	Johnny Depp		The Beach	Babel	Inception	Leo_Caprio
Man of Steel	Henry Cavill		Revolutionary Road	By the Sea	The Revenant	Leo_Caprio
The Chronicles of Narnia: The Lion, the Witch and the Wardrobe	Peter Dinklage		The Man in the Iron Mask	Killing Them Softly	The Aviator	Leo_Caprio
The Avengers	Chris Hemsworth		J. Edgar	True Romance	Django Unchained	Leo_Caprio
Pirates of the Caribbean: On Stranger Tides	Johnny Depp		The Quick and the Dead		Blood Diamond	Leo_Caprio
Men in Black 3	Will Smith		Marvin's Room		The Wolf of Wall Street	Leo_Caprio
The Hobbit: The Battle of the Five Armies	Aidan Turner		Romeo + Juliet		Gangs of New York	Leo_Caprio
The Amazing Spider-Man	Emma Stone		The Great Gatsby		The Departed	Leo_Caprio
Robin Hood	Mark Addy				Shutter Island	Leo_Caprio
The Hobbit: The Desolation of Smaug	Aidan Turner				Body of Lies	Leo_Caprio
The Golden Compass	Christopher Lee				Catch Me If You Can	Leo_Caprio
King Kong	Naomi Watts				The Beach	Leo_Caprio
Titanic	Leonardo DiCaprio				Revolutionary Road	Leo_Caprio
Captain America: Civil War	Robert Downey Jr.				The Man in the Iron Mask	Leo_Caprio
Battleship	Liam Neeson				J. Edgar	Leo_Caprio
Jurassic World	Bryce Dallas Howard				The Quick and the Dead	Leo_Caprio
Skyfall	Albert Finney				Marvin's Room	Leo_Caprio
Spider-Man 2	J.K. Simmons				Romeo + Juliet	Leo_Caprio
Iron Man 3	Robert Downey Jr.				The Great Gatsby	Leo_Caprio
Alice in Wonderland	Johnny Depp				The Curious Case of Benjamin Button	Brad_Pitt
X-Men: The Last Stand	Hugh Jackman				Troy	Brad_Pitt
Monsters University	Steve Buscemi				Ocean's Twelve	Brad_Pitt
Transformers: Revenge of the Fallen	Glenn Morshower				Mr. & Mrs. Smith	Brad_Pitt
Transformers: Age of Extinction	Bing Li				Spy Game	Brad_Pitt
Oz the Great and Powerful	Tim Holmes				Ocean's Eleven	Brad_Pitt

Find the mean of the num_critics_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.

