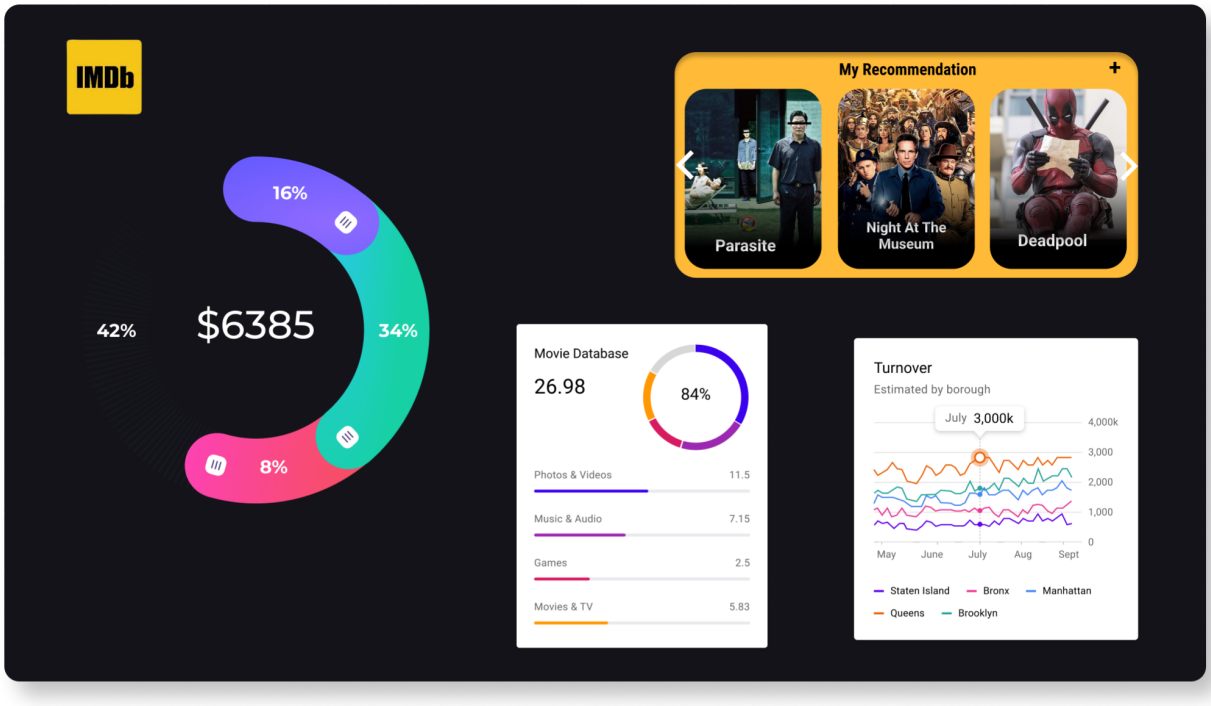
IMDB Movie Analysis



By Aayushi Singh

Project Description

You have a dataset having various columns of different IMDB Movies. Frame the problem. For this task, you will need to define a problem you want to shed some light on.

Approach

We analysis the given data by first cleaning the data then according to the give problem statement we ask questions by “5 whys” it is called root cause analysis

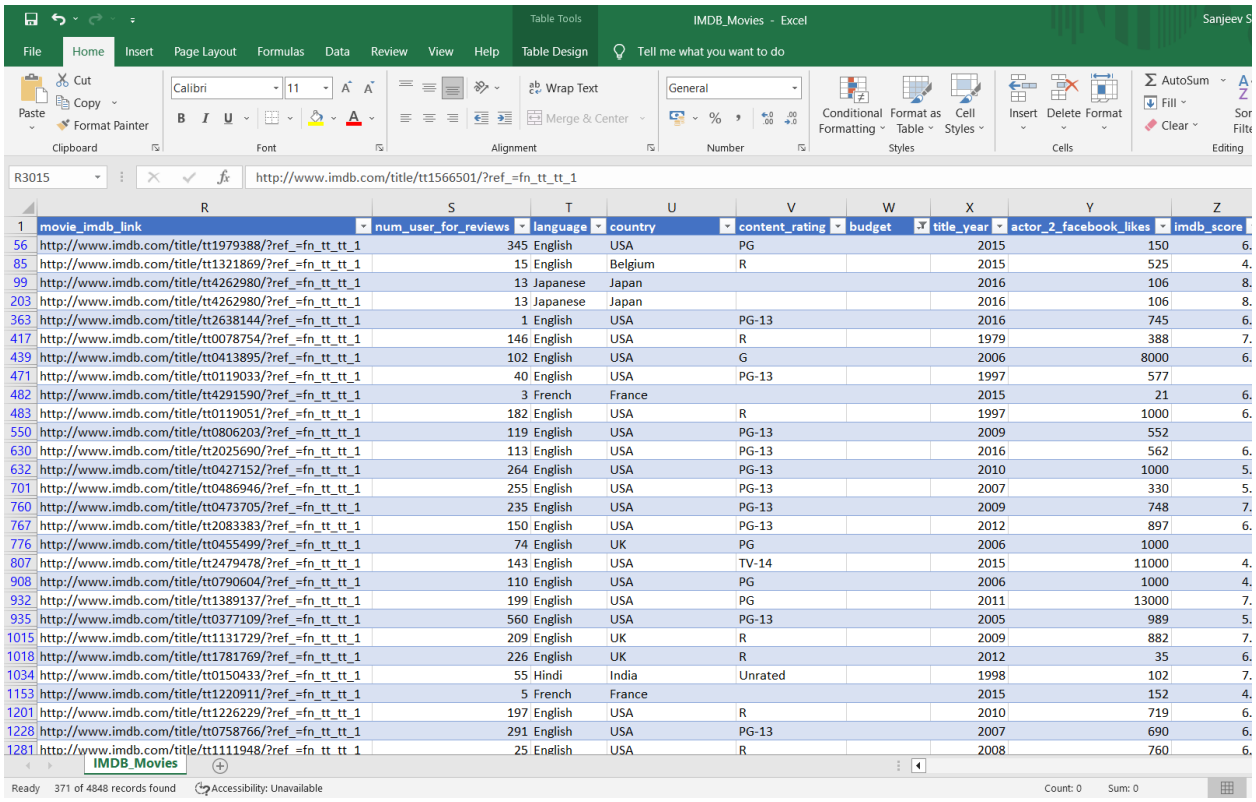
Tech Stack Used

Ms excel
Google Docs
Google Sheets

Results and Insights

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

Insights: You clean the data by clearing out the blanks



IMDB_Movies - Excel

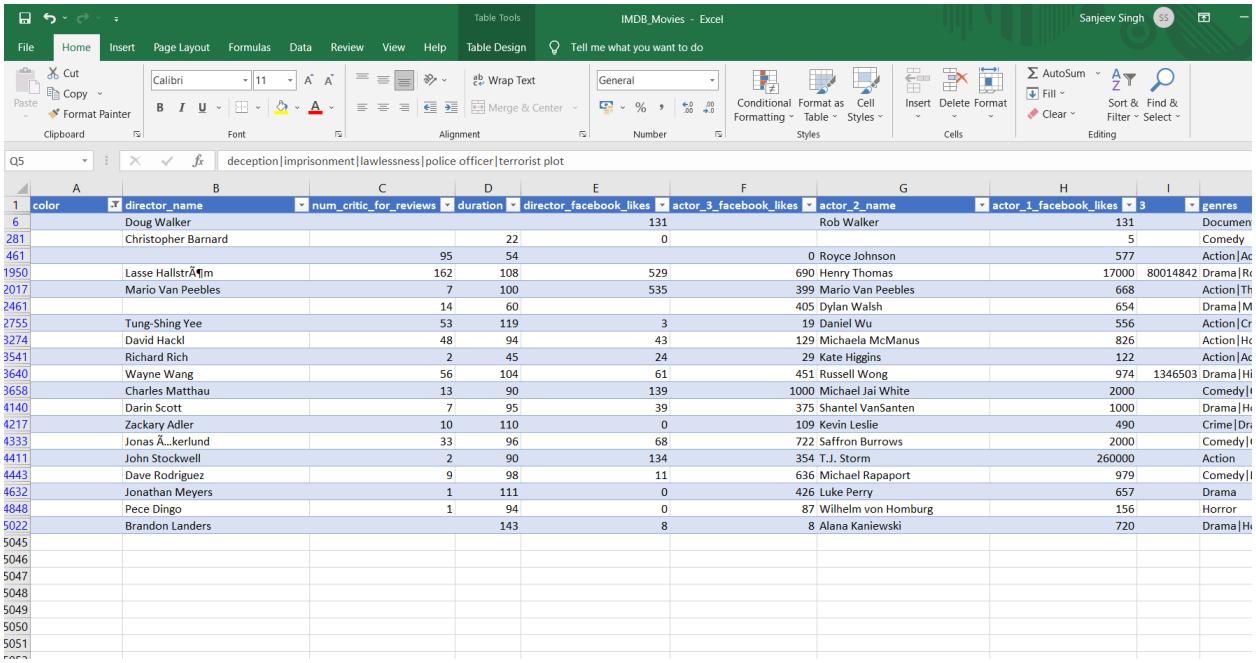
R3015 | http://www.imdb.com/title/tt1566501/?ref_=fn_tt_tt_1

	R	S	T	U	V	W	X	Y	Z
1	movie_imdb_link	num_user_for_reviews	language	country	content_rating	budget	title_year	actor_2_facebook_likes	imdb_score
56	http://www.imdb.com/title/tt1979388/?ref_=fn_tt_tt_1	345	English	USA	PG		2015	150	6.
85	http://www.imdb.com/title/tt1321869/?ref_=fn_tt_tt_1	15	English	Belgium	R		2015	525	4.
99	http://www.imdb.com/title/tt4262980/?ref_=fn_tt_tt_1	13	Japanese	Japan			2016	106	8.
203	http://www.imdb.com/title/tt4262980/?ref_=fn_tt_tt_1	13	Japanese	Japan			2016	106	8.
363	http://www.imdb.com/title/tt2638144/?ref_=fn_tt_tt_1	1	English	USA	PG-13		2016	745	6.
417	http://www.imdb.com/title/tt0078754/?ref_=fn_tt_tt_1	146	English	USA	R		1979	388	7.
439	http://www.imdb.com/title/tt0413895/?ref_=fn_tt_tt_1	102	English	USA	G		2006	8000	6.
471	http://www.imdb.com/title/tt0119033/?ref_=fn_tt_tt_1	40	English	USA	PG-13		1997	577	
482	http://www.imdb.com/title/tt4291590/?ref_=fn_tt_tt_1	3	French	France			2015	21	6.
483	http://www.imdb.com/title/tt0119051/?ref_=fn_tt_tt_1	182	English	USA	R		1997	1000	6.
550	http://www.imdb.com/title/tt0806203/?ref_=fn_tt_tt_1	119	English	USA	PG-13		2009	552	
630	http://www.imdb.com/title/tt2025690/?ref_=fn_tt_tt_1	113	English	USA	PG-13		2016	562	6.
632	http://www.imdb.com/title/tt0427152/?ref_=fn_tt_tt_1	264	English	USA	PG-13		2010	1000	5.
701	http://www.imdb.com/title/tt0486946/?ref_=fn_tt_tt_1	255	English	USA	PG-13		2007	330	5.
760	http://www.imdb.com/title/tt0473705/?ref_=fn_tt_tt_1	235	English	USA	PG-13		2009	748	7.
767	http://www.imdb.com/title/tt2083383/?ref_=fn_tt_tt_1	150	English	USA	PG-13		2012	897	6.
776	http://www.imdb.com/title/tt0455499/?ref_=fn_tt_tt_1	74	English	UK	PG		2006	1000	
807	http://www.imdb.com/title/tt2479478/?ref_=fn_tt_tt_1	143	English	USA	TV-14		2015	11000	4.
908	http://www.imdb.com/title/tt0790604/?ref_=fn_tt_tt_1	110	English	USA	PG		2006	1000	4.
932	http://www.imdb.com/title/tt1389137/?ref_=fn_tt_tt_1	199	English	USA	PG		2011	13000	7.
935	http://www.imdb.com/title/tt0377109/?ref_=fn_tt_tt_1	560	English	USA	PG-13		2005	989	5.
1015	http://www.imdb.com/title/tt1131729/?ref_=fn_tt_tt_1	209	English	UK	R		2009	882	7.
1018	http://www.imdb.com/title/tt1781769/?ref_=fn_tt_tt_1	226	English	UK	R		2012	35	6.
1034	http://www.imdb.com/title/tt0150433/?ref_=fn_tt_tt_1	55	Hindi	India	Unrated		1998	102	7.
1153	http://www.imdb.com/title/tt1220911/?ref_=fn_tt_tt_1	5	French	France			2015	152	4.
1201	http://www.imdb.com/title/tt1226229/?ref_=fn_tt_tt_1	197	English	USA	R		2010	719	6.
1228	http://www.imdb.com/title/tt0758766/?ref_=fn_tt_tt_1	291	English	USA	PG-13		2007	690	6.
1281	http://www.imdb.com/title/tt1111948/?ref_=fn_tt_tt_1	25	English	USA	R		2008	760	6.

IMDB_Movies

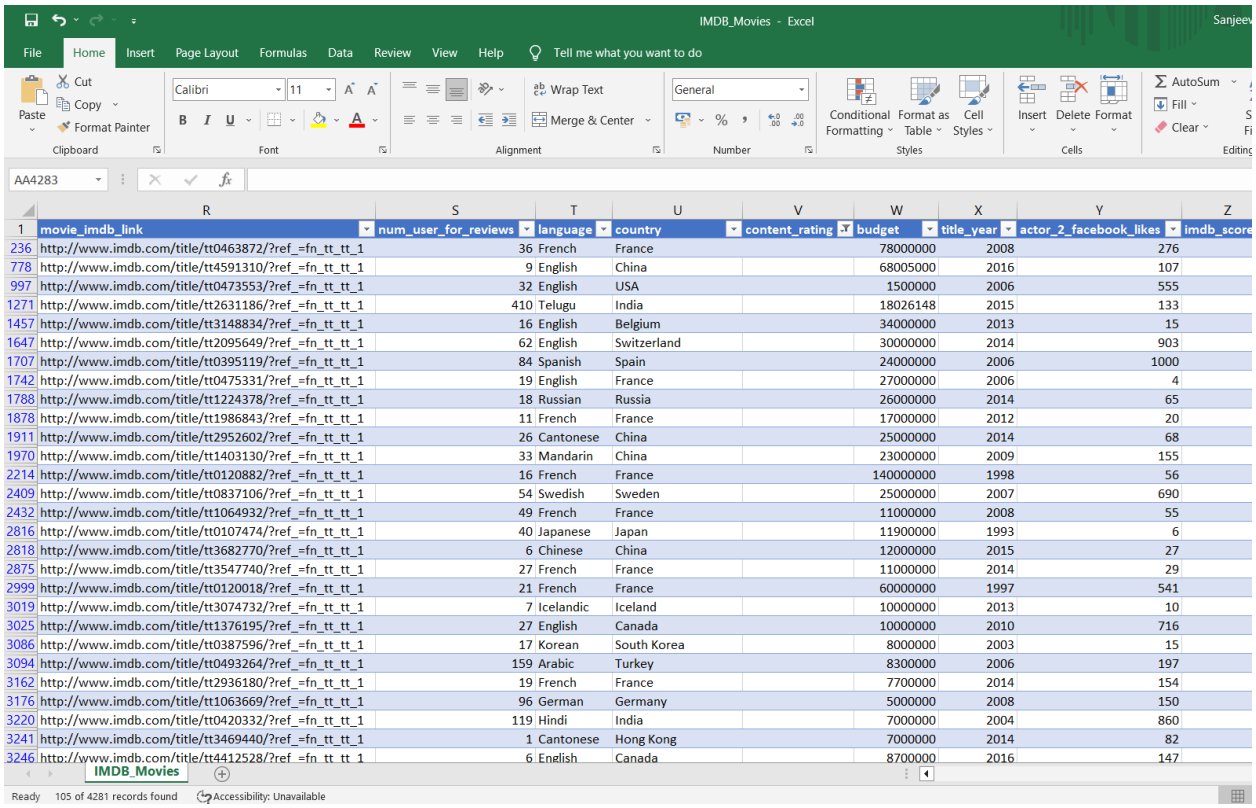
Ready 371 of 4848 records found Accessibility: Unavailable Count: 0 Sum: 0

Blanks in budget



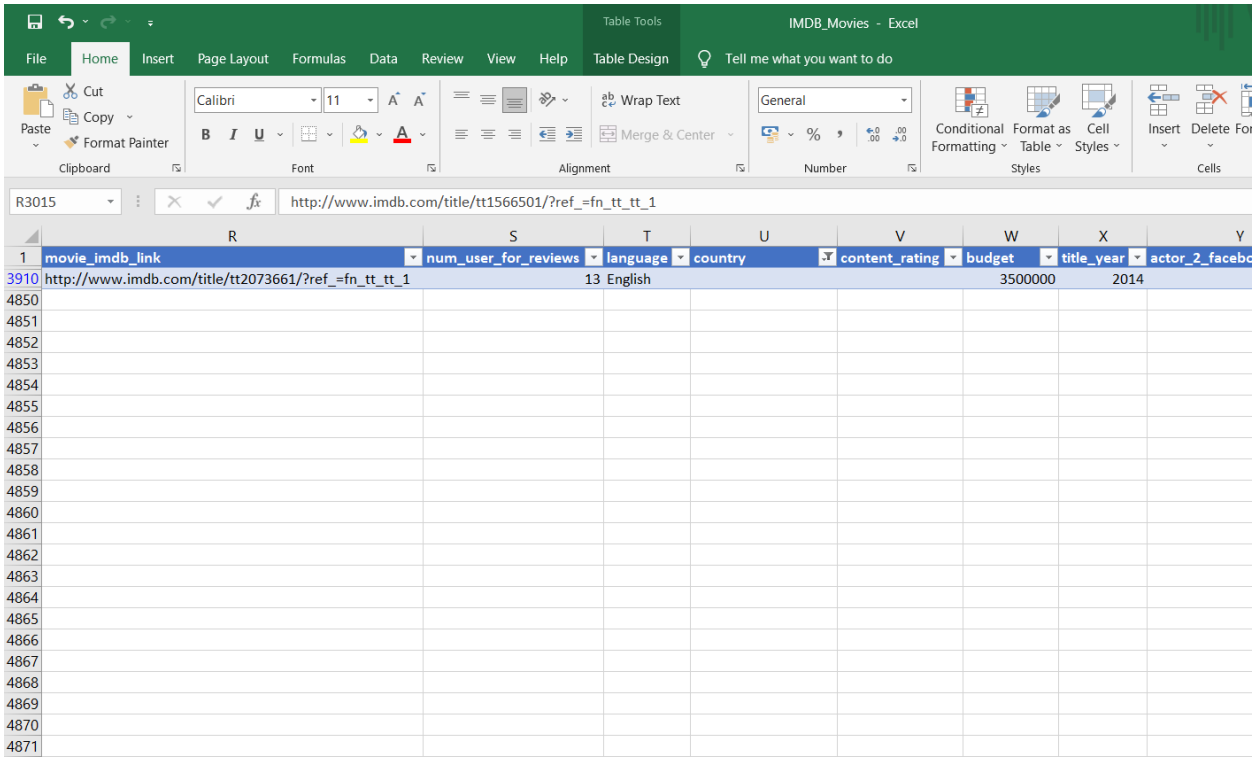
	A	B	C	D	E	F	G	H	I	
1	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	3	genres
6		Doug Walker			131		Rob Walker	131		Documen
281		Christopher Barnard		22	0			5		Comedy
461			95	54		0	Royce Johnson	577		Action|Ac
1950		Lasse HallstrÃ¶m	162	108	529	690	Henry Thomas	17000	80014842	Drama|Ro
2017		Mario Van Peebles	7	100	535	399	Mario Van Peebles	668		Action|Th
2461			14	60		405	Dylan Walsh	654		Drama|M
2755		Tung-Shing Yee	53	119	3	19	Daniel Wu	556		Action|Cr
3274		David Hackl	48	94	43	129	Michaela McManus	826		Action|Ho
3541		Richard Rich	2	45	24	29	Kate Higgins	122		Action|Ac
3640		Wayne Wang	56	104	61	451	Russell Wong	974	1346503	Drama|Hi
3658		Charles Matthau	13	90	139	1000	Michael Jai White	2000		Comedy|C
4140		Darin Scott	7	95	39	375	Shantel VanSanten	1000		Drama|Ho
4217		Zackary Adler	10	110	0	109	Kevin Leslie	490		Crime|Dra
4333		Jonas Ã…kerlund	33	96	68	722	Saffron Burrows	2000		Comedy|C
4411		John Stockwell	2	90	134	354	T.J. Storm	260000		Action
4443		Dave Rodriguez	9	98	11	636	Michael Rapaport	979		Comedy|I
4632		Jonathan Meyers	1	111	0	426	Luke Perry	657		Drama
4848		Pece Dingo	1	94	0	87	Wilhelm von Homburg	156		Horror
5022		Brandon Landers		143	8	8	Alana Kaniewski	720		Drama|Ho
5045										
5046										
5047										
5048										
5049										
5050										
5051										

Blanks in color



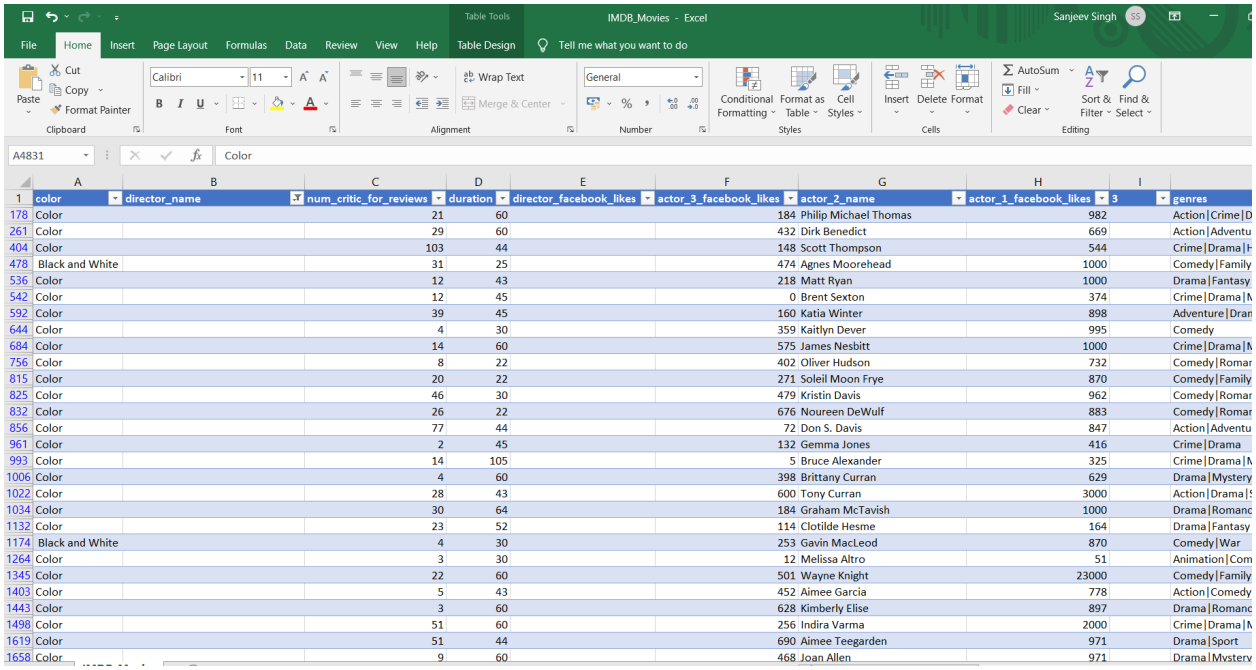
	R	S	T	U	V	W	X	Y	Z
1	movie_imdb_link	num_user_for_reviews	language	country	content_rating	budget	title_year	actor_2_facebook_likes	imdb_score
236	http://www.imdb.com/title/tt0463872/?ref_=fn_tt_tt_1	36	French	France		78000000	2008	276	
778	http://www.imdb.com/title/tt4591310/?ref_=fn_tt_tt_1	9	English	China		68005000	2016	107	
997	http://www.imdb.com/title/tt0473553/?ref_=fn_tt_tt_1	32	English	USA		1500000	2006	555	
1271	http://www.imdb.com/title/tt2631186/?ref_=fn_tt_tt_1	410	Telugu	India		18026148	2015	133	
1457	http://www.imdb.com/title/tt3148834/?ref_=fn_tt_tt_1	16	English	Belgium		34000000	2013	15	
1647	http://www.imdb.com/title/tt2095649/?ref_=fn_tt_tt_1	62	English	Switzerland		30000000	2014	903	
1707	http://www.imdb.com/title/tt0395119/?ref_=fn_tt_tt_1	84	Spanish	Spain		24000000	2006	1000	
1742	http://www.imdb.com/title/tt0475331/?ref_=fn_tt_tt_1	19	English	France		27000000	2006	4	
1788	http://www.imdb.com/title/tt1224378/?ref_=fn_tt_tt_1	18	Russian	Russia		26000000	2014	65	
1878	http://www.imdb.com/title/tt1986843/?ref_=fn_tt_tt_1	11	French	France		17000000	2012	20	
1911	http://www.imdb.com/title/tt2952602/?ref_=fn_tt_tt_1	26	Cantonese	China		25000000	2014	68	
1970	http://www.imdb.com/title/tt1403130/?ref_=fn_tt_tt_1	33	Mandarin	China		23000000	2009	155	
2214	http://www.imdb.com/title/tt0120882/?ref_=fn_tt_tt_1	16	French	France		140000000	1998	56	
2409	http://www.imdb.com/title/tt0837106/?ref_=fn_tt_tt_1	54	Swedish	Sweden		25000000	2007	690	
2432	http://www.imdb.com/title/tt1064932/?ref_=fn_tt_tt_1	49	French	France		11000000	2008	55	
2816	http://www.imdb.com/title/tt0107474/?ref_=fn_tt_tt_1	40	Japanese	Japan		11900000	1993	6	
2818	http://www.imdb.com/title/tt3682770/?ref_=fn_tt_tt_1	6	Chinese	China		12000000	2015	27	
2875	http://www.imdb.com/title/tt3547740/?ref_=fn_tt_tt_1	27	French	France		11000000	2014	29	
2999	http://www.imdb.com/title/tt0120018/?ref_=fn_tt_tt_1	21	French	France		60000000	1997	541	
3019	http://www.imdb.com/title/tt3074732/?ref_=fn_tt_tt_1	7	Icelandic	Iceland		10000000	2013	10	
3025	http://www.imdb.com/title/tt1376195/?ref_=fn_tt_tt_1	27	English	Canada		10000000	2010	716	
3086	http://www.imdb.com/title/tt0387596/?ref_=fn_tt_tt_1	17	Korean	South Korea		8000000	2003	15	
3094	http://www.imdb.com/title/tt0493264/?ref_=fn_tt_tt_1	159	Arabic	Turkey		8300000	2006	197	
3162	http://www.imdb.com/title/tt2936180/?ref_=fn_tt_tt_1	19	French	France		7700000	2014	154	
3176	http://www.imdb.com/title/tt1063669/?ref_=fn_tt_tt_1	96	German	Germany		5000000	2008	150	
3220	http://www.imdb.com/title/tt0420332/?ref_=fn_tt_tt_1	119	Hindi	India		7000000	2004	860	
3241	http://www.imdb.com/title/tt3469440/?ref_=fn_tt_tt_1	1	Cantonese	Hong Kong		7000000	2014	82	
3246	http://www.imdb.com/title/tt4412528/?ref_=fn_tt_tt_1	6	English	Canada		8700000	2016	147	

Blanks in Content rating



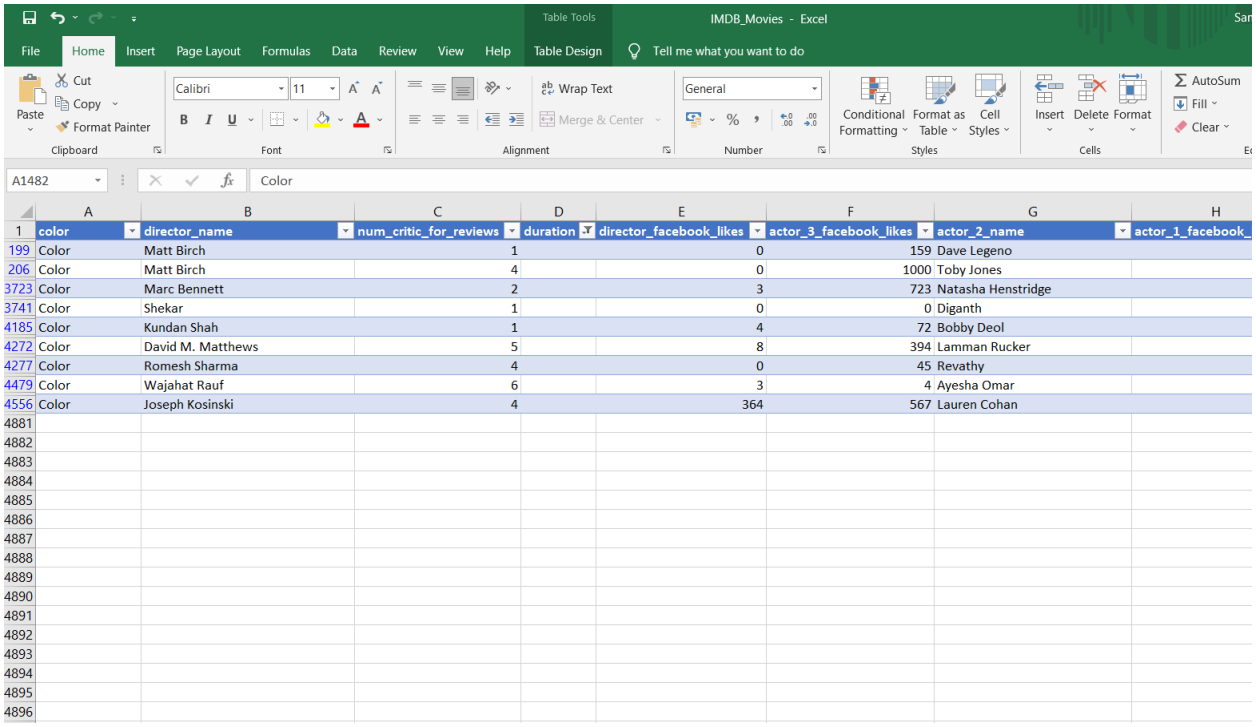
	R	S	T	U	V	W	X	Y
1	movie_imdb_link	num_user_for_reviews	language	country	content_rating	budget	title_year	actor_2_facebo
3910	http://www.imdb.com/title/tt2073661/?ref_=fn_tt_tt_1	13	English			3500000	2014	
4850								
4851								
4852								
4853								
4854								
4855								
4856								
4857								
4858								
4859								
4860								
4861								
4862								
4863								
4864								
4865								
4866								
4867								
4868								
4869								
4870								
4871								

Blanks in country



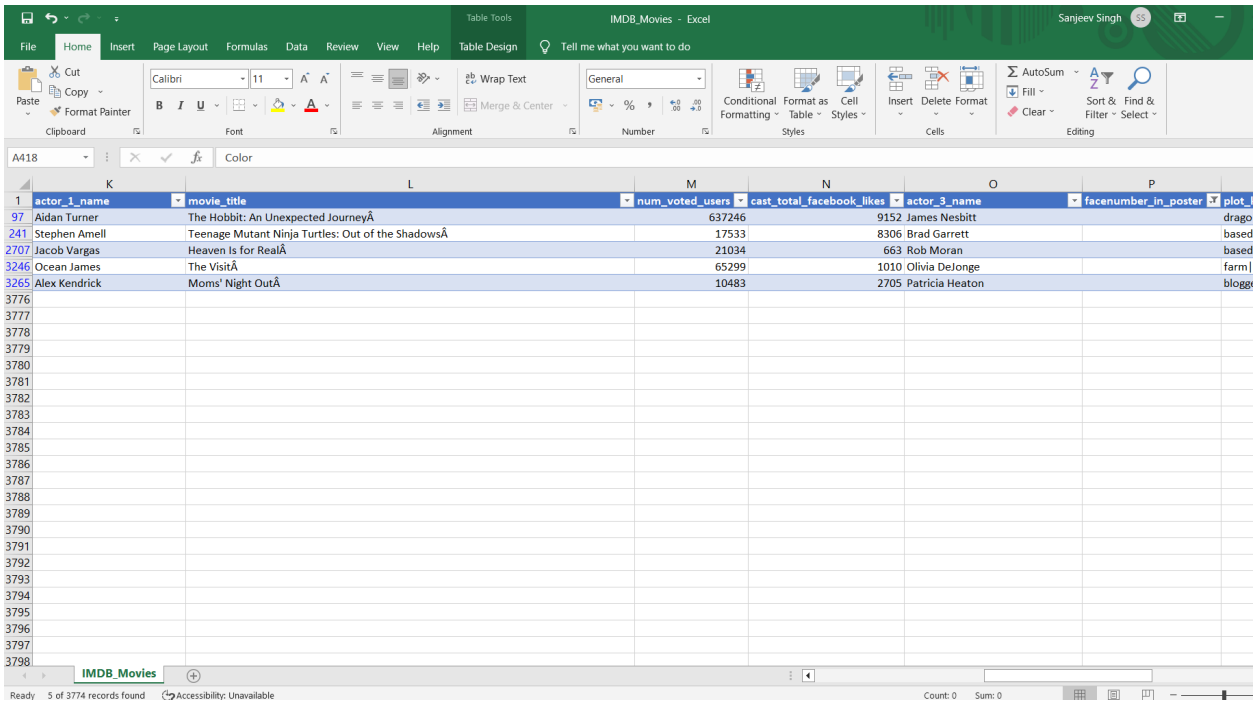
	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	3	genres
178	Color		21	60		184	Philip Michael Thomas	982		Action|Crime|D
261	Color		29	60		432	Dirk Benedict	669		Action|Adventu
404	Color		103	44		148	Scott Thompson	544		Crime|Drama|H
478	Black and White		31	25		474	Agnes Moorehead	1000		Comedy|Family
536	Color		12	43		218	Matt Ryan	1000		Drama|Fantasy
542	Color		12	45		0	Brent Sexton	374		Crime|Drama|M
592	Color		39	45		160	Katia Winter	898		Adventure|Dran
644	Color		4	30		359	Kaitlyn Dever	995		Comedy
684	Color		14	60		575	James Nesbitt	1000		Crime|Drama|M
756	Color		8	22		402	Oliver Hudson	732		Comedy|Roman
815	Color		20	22		271	Soleil Moon Frye	870		Comedy|Family
825	Color		46	30		479	Kristin Davis	962		Comedy|Roman
832	Color		26	22		676	Noureen DeWulf	883		Comedy|Roman
856	Color		77	44		72	Don S. Davis	847		Action|Adventu
961	Color		2	45		132	Gemma Jones	416		Crime|Drama
993	Color		14	105		5	Bruce Alexander	325		Crime|Drama|M
1006	Color		4	60		398	Brittany Curran	629		Drama|Mystery
1022	Color		28	43		600	Tony Curran	3000		Action|Drama|S
1034	Color		30	64		184	Graham McTavish	1000		Drama|Romanc
1132	Color		23	52		114	Clotilde Hesme	164		Drama|Fantasy
1174	Black and White		4	30		253	Gavin MacLeod	870		Comedy|War
1264	Color		3	30		12	Melissa Altro	51		Animation|Com
1345	Color		22	60		501	Wayne Knight	23000		Comedy|Family
1403	Color		5	43		452	Aimee Garcia	778		Action|Comedy
1443	Color		3	60		628	Kimberly Elise	897		Drama|Romanc
1498	Color		51	60		256	Indira Varma	2000		Crime|Drama|M
1619	Color		51	44		690	Aimee Teegarden	971		Drama|Sport
1658	Color		9	60		468	Joan Allen	971		Drama|Mystery

Blanks in director name



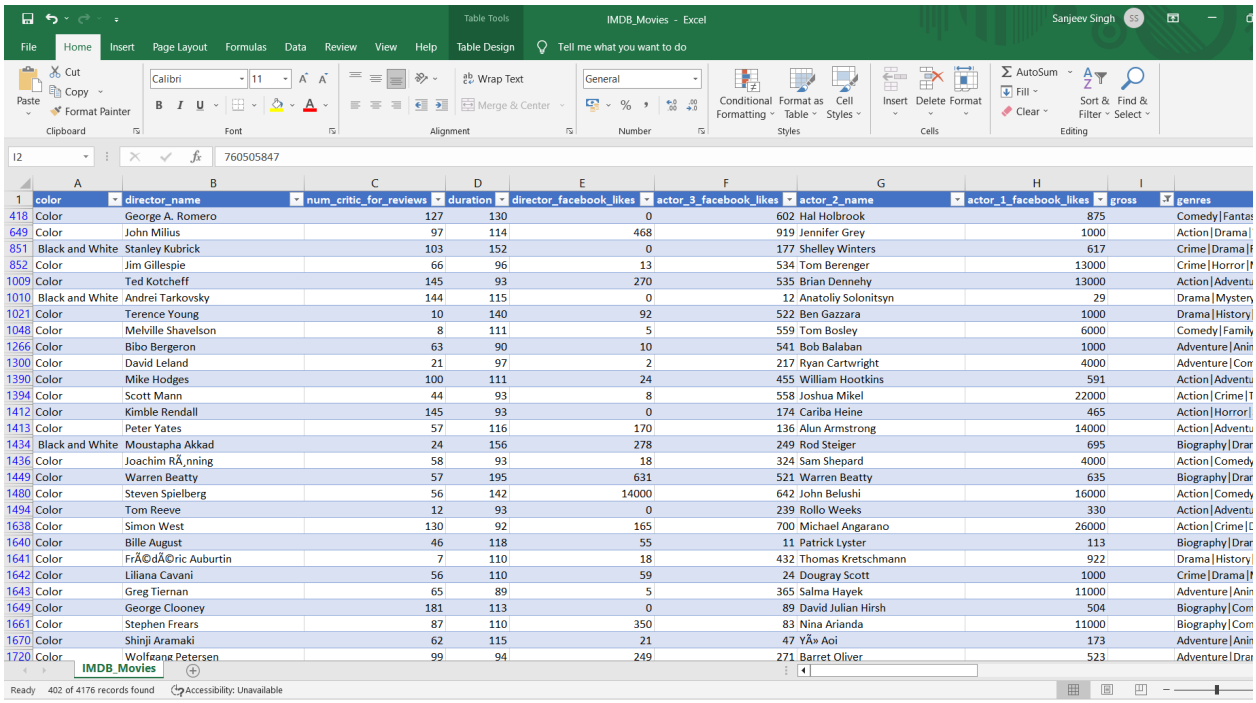
	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_
199	Color	Matt Birch	1		0	159	Dave Legeno	
206	Color	Matt Birch	4		0	1000	Toby Jones	
3723	Color	Marc Bennett	2		3	723	Natasha Henstridge	
3741	Color	Shekar	1		0	0	Diganth	
4185	Color	Kundan Shah	1		4	72	Bobby Deol	
4272	Color	David M. Matthews	5		8	394	Lamman Rucker	
4277	Color	Romesh Sharma	4		0	45	Revathy	
4479	Color	Wajahat Rauf	6		3	4	Ayesha Omar	
4556	Color	Joseph Kosinski	4		364	567	Lauren Cohan	

Blanks in duration



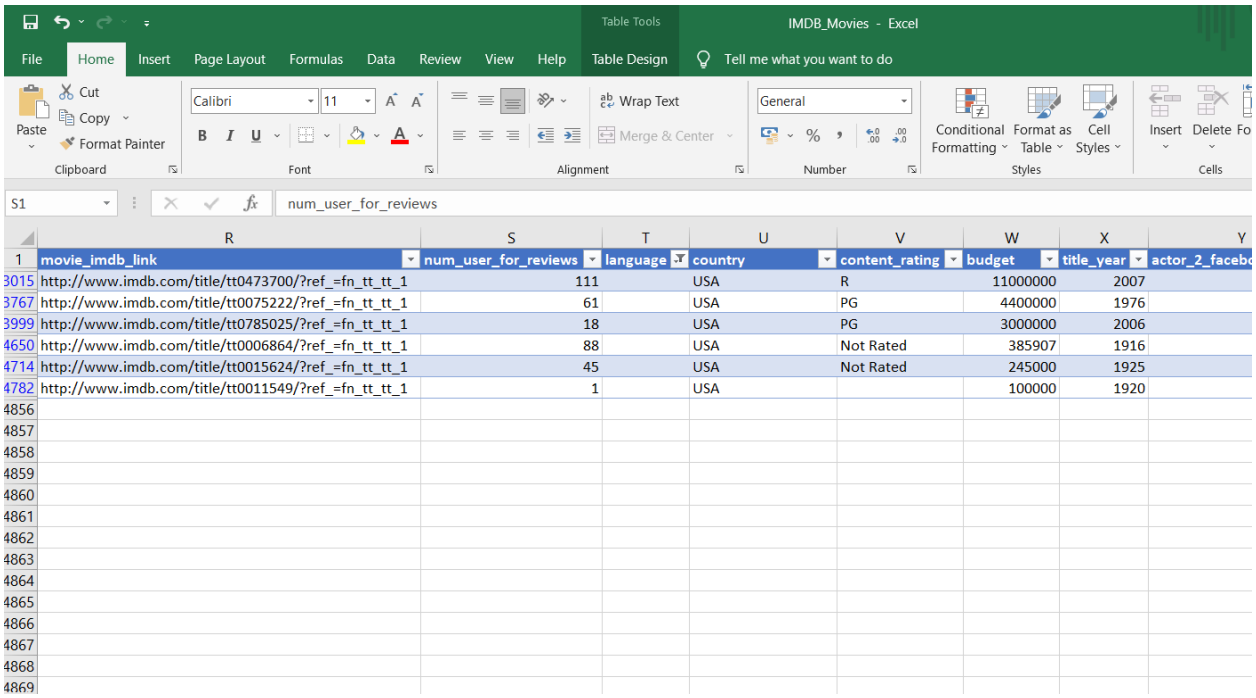
	K	L	M	N	O	P
1	actor_1_name	movie_title	num_voted_users	cast_total_facebook_likes	actor_3_name	facenumber_in_poster
97	Aidan Turner	The Hobbit: An Unexpected JourneyÂ	637246	9152	James Nesbitt	
241	Stephen Amell	Teenage Mutant Ninja Turtles: Out of the ShadowsÂ	17533	8306	Brad Garrett	
2707	Jacob Vargas	Heaven Is for RealÂ	21034	663	Rob Moran	
3246	Ocean James	The VisitÂ	65299	1010	Olivia DeJonge	
3265	Alex Kendrick	Moms' Night OutÂ	10483	2705	Patricia Heaton	

Blanks in facenumber in posters



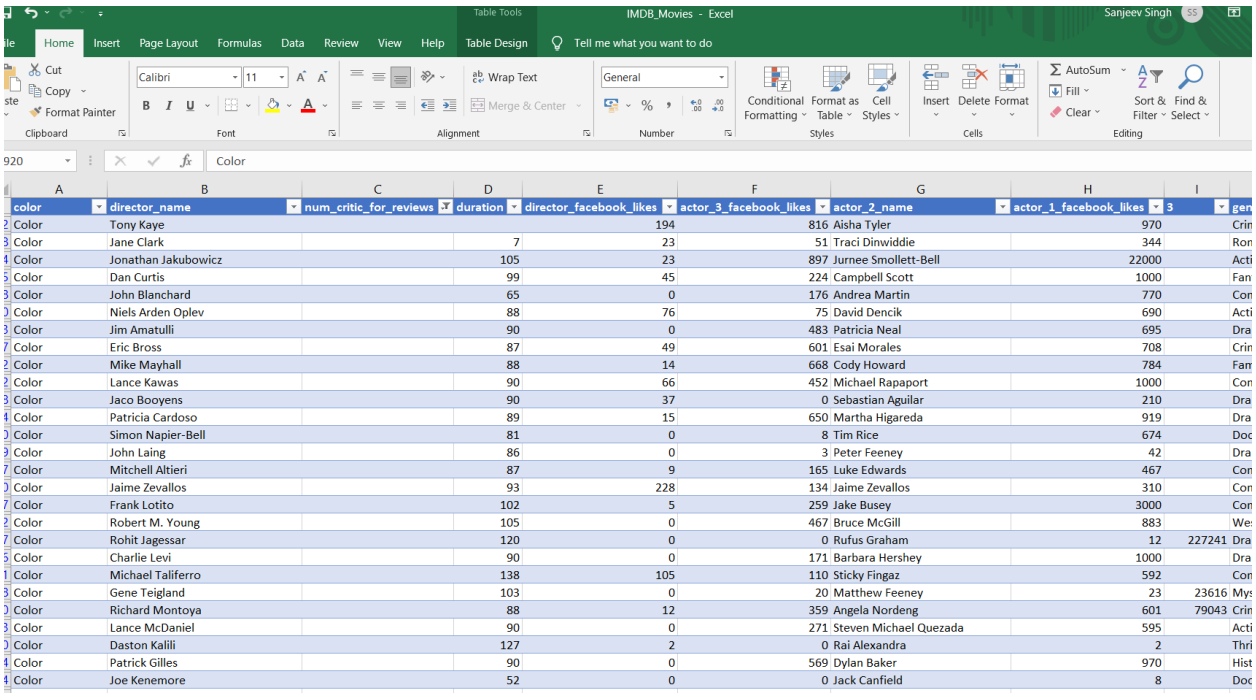
	A	B	C	D	E	F	G	H	I
1	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross
418	Color	George A. Romero	127	130	0	602	Hal Holbrook	875	
649	Color	John Milius	97	114	468	919	Jennifer Grey	1000	
851	Black and White	Stanley Kubrick	103	152	0	177	Shelley Winters	617	
852	Color	Jim Gillespie	66	96	13	534	Tom Berenger	13000	
1009	Color	Ted Kotcheff	145	93	270	535	Brian Dennehy	13000	
1010	Black and White	Andrei Tarkovsky	144	115	0	12	Anatoliy Solonitsyn	29	
1021	Color	Terence Young	10	140	92	522	Ben Gazzara	1000	
1048	Color	Melville Shavelson	8	111	5	559	Tom Bosley	6000	
1266	Color	Bibo Bergeron	63	90	10	541	Bob Balaban	1000	
1300	Color	David Leland	21	97	2	217	Ryan Cartwright	4000	
1390	Color	Mike Hodges	100	111	24	455	William Hootkins	591	
1394	Color	Scott Mann	44	93	8	558	Joshua Mikel	22000	
1412	Color	Kimble Rendall	145	93	0	174	Cariba Heine	465	
1413	Color	Peter Yates	57	116	170	136	Alun Armstrong	14000	
1434	Black and White	Moustapha Akkad	24	156	278	249	Rod Steiger	695	
1436	Color	Joachim RÃ¸nning	58	93	18	324	Sam Shepard	4000	
1449	Color	Warren Beatty	57	195	631	521	Warren Beatty	635	
1480	Color	Steven Spielberg	56	142	14000	642	John Belushi	16000	
1494	Color	Tom Reeve	12	93	0	239	Rollo Weeks	330	
1638	Color	Simon West	130	92	165	700	Michael Angarano	26000	
1640	Color	Bille August	46	118	55	11	Patrick Lyster	113	
1641	Color	FrÃ©dÃ©ric Auburtin	7	110	18	432	Thomas Kretschmann	922	
1642	Color	Liliana Cavani	56	110	59	24	Dougray Scott	1000	
1643	Color	Greg Tiernan	65	89	5	365	Salma Hayek	11000	
1649	Color	George Clooney	181	113	0	89	David Julian Hirsh	504	
1661	Color	Stephen Frears	87	110	350	83	Nina Arianda	11000	
1670	Color	Shinji Aramaki	62	115	21	47	YÃ» Aoi	173	
1720	Color	Wolfgang Petersen	99	94	249	271	Barret Oliver	523	

Blanks in gross



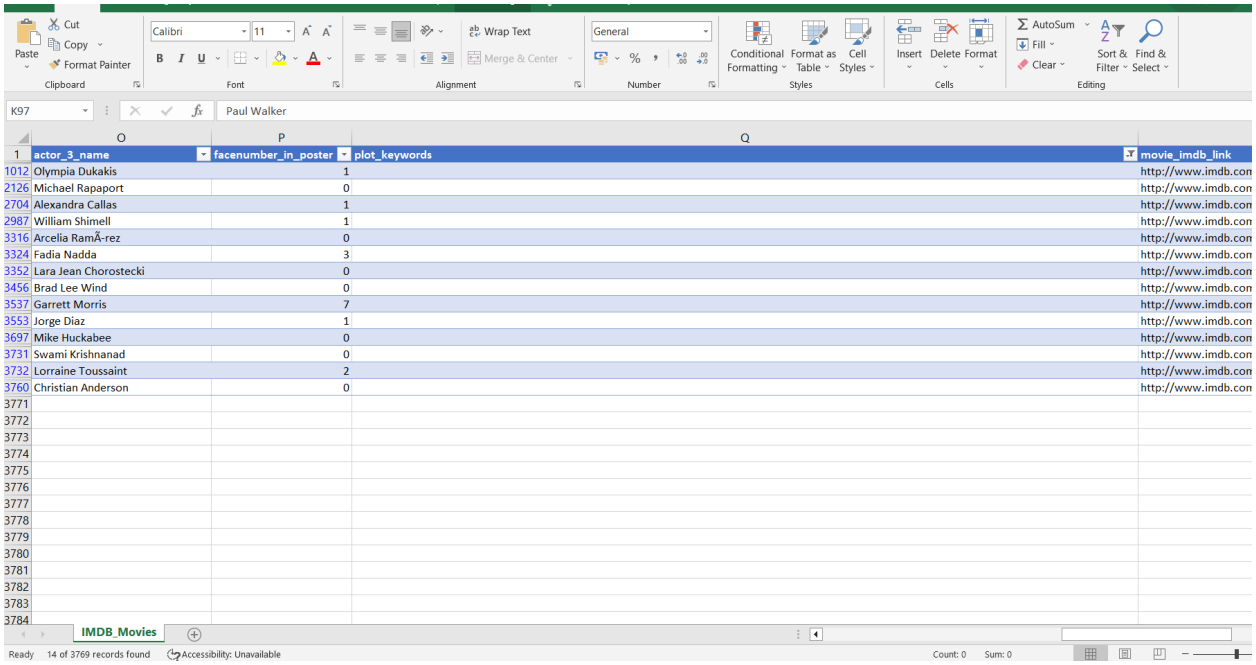
	R	S	T	U	V	W	X	Y
1	movie_imdb_link	num_user_for_reviews	language	country	content_rating	budget	title_year	actor_2_faceb
3015	http://www.imdb.com/title/tt0473700/?ref_=fn_tt_tt_1	111		USA	R	11000000	2007	
3767	http://www.imdb.com/title/tt0075222/?ref_=fn_tt_tt_1	61		USA	PG	4400000	1976	
3999	http://www.imdb.com/title/tt0785025/?ref_=fn_tt_tt_1	18		USA	PG	3000000	2006	
4650	http://www.imdb.com/title/tt0006864/?ref_=fn_tt_tt_1	88		USA	Not Rated	385907	1916	
4714	http://www.imdb.com/title/tt0015624/?ref_=fn_tt_tt_1	45		USA	Not Rated	245000	1925	
4782	http://www.imdb.com/title/tt0011549/?ref_=fn_tt_tt_1	1		USA		100000	1920	
4856								
4857								
4858								
4859								
4860								
4861								
4862								
4863								
4864								
4865								
4866								
4867								
4868								
4869								

Blanks in language



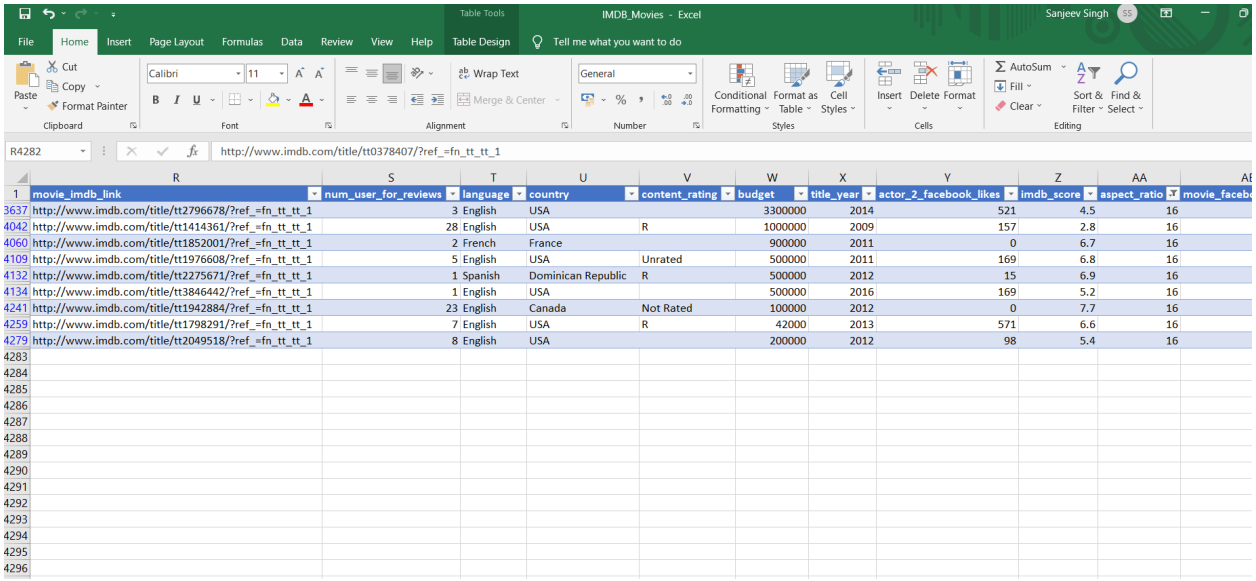
	A	B	C	D	E	F	G	H	I	
	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	3	ger
2	Color	Tony Kaye			194	816	Aisha Tyler	970		Crin
3	Color	Jane Clark		7	23	51	Traci Dinwiddie	344		Ron
4	Color	Jonathan Jakubowicz		105	23	897	Jurnee Smollett-Bell	22000		Acti
5	Color	Dan Curtis		99	45	224	Campbell Scott	1000		Fan
3	Color	John Blanchard		65	0	176	Andrea Martin	770		Con
0	Color	Niels Arden Oplev		88	76	75	David Dencik	690		Acti
3	Color	Jim Amatulli		90	0	483	Patricia Neal	695		Dra
7	Color	Eric Bross		87	49	601	Esai Morales	708		Crin
2	Color	Mike Mayhall		88	14	668	Cody Howard	784		Fan
2	Color	Lance Kawas		90	66	452	Michael Rapaport	1000		Con
3	Color	Jaco Booyens		90	37	0	Sebastian Aguilar	210		Dra
1	Color	Patricia Cardoso		89	15	650	Martha Higareda	919		Dra
0	Color	Simon Napier-Bell		81	0	8	Tim Rice	674		Doc
9	Color	John Laing		86	0	3	Peter Feeney	42		Dra
7	Color	Mitchell Altieri		87	9	165	Luke Edwards	467		Con
0	Color	Jaime Zevallos		93	228	134	Jaime Zevallos	310		Con
7	Color	Frank Lotito		102	5	259	Jake Busey	3000		Con
2	Color	Robert M. Young		105	0	467	Bruce McGill	883		We
7	Color	Rohit Jagessar		120	0	0	Rufus Graham	12	227241	Dra
5	Color	Charlie Levi		90	0	171	Barbara Hershey	1000		Dra
1	Color	Michael Taliferro		138	105	110	Sticky Fingaz	592		Con
3	Color	Gene Teigland		103	0	20	Matthew Feeney	23	23616	Mys
0	Color	Richard Montoya		88	12	359	Angela Nordeng	601	79043	Crin
3	Color	Lance McDaniel		90	0	271	Steven Michael Quezada	595		Acti
0	Color	Daston Kalili		127	2	0	Rai Alexandra	2		Thri
4	Color	Patrick Gilles		90	0	569	Dylan Baker	970		Hist
4	Color	Joe Kenemore		52	0	0	Jack Canfield	8		Doc

Blanks in number of critics for review



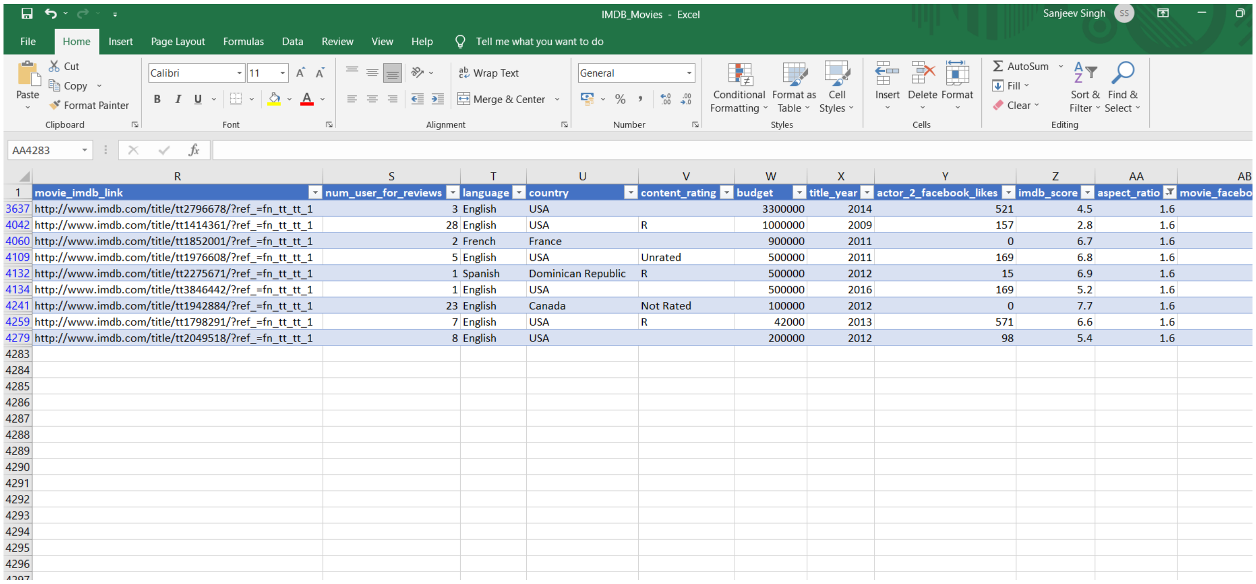
	O	P	Q	
1	actor_3_name	facenumber_in_poster	plot_keywords	movie_imdb_link
1012	Olympia Dukakis	1		http://www.imdb.com
2126	Michael Rapaport	0		http://www.imdb.com
2704	Alexandra Callas	1		http://www.imdb.com
2987	William Shimell	1		http://www.imdb.com
3316	Arcelia RamÃ-rez	0		http://www.imdb.com
3324	Fadia Nadda	3		http://www.imdb.com
3352	Lara Jean Chorostecki	0		http://www.imdb.com
3456	Brad Lee Wind	0		http://www.imdb.com
3537	Garrett Morris	7		http://www.imdb.com
3553	Jorge Diaz	1		http://www.imdb.com
3697	Mike Huckabee	0		http://www.imdb.com
3731	Swami Krishnanad	0		http://www.imdb.com
3732	Lorraine Toussaint	2		http://www.imdb.com
3760	Christian Anderson	0		http://www.imdb.com

Blanks in plot key words



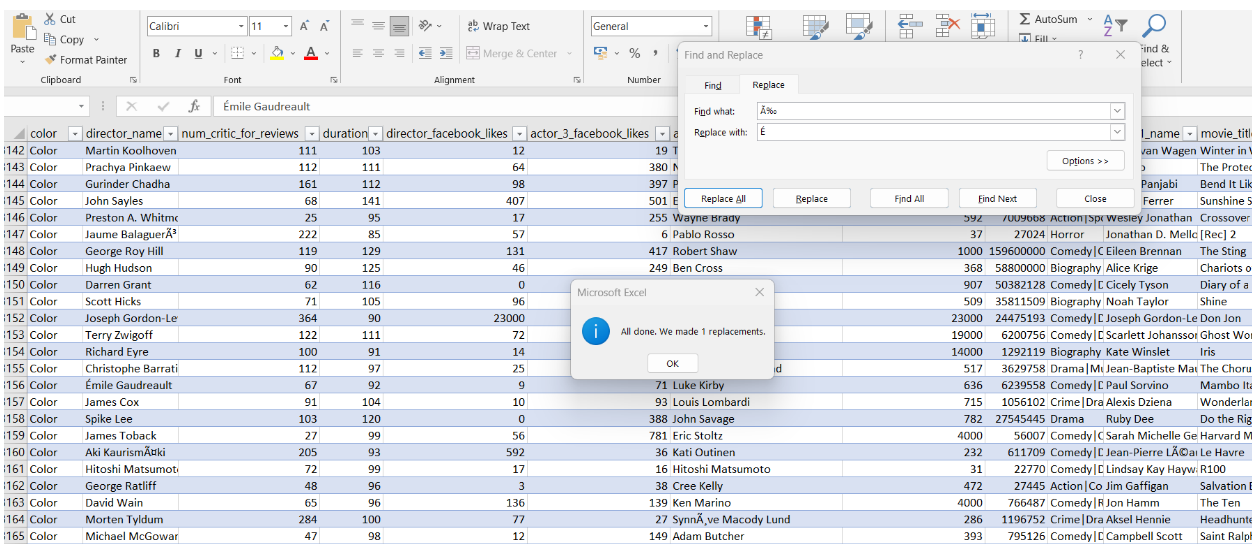
	R	S	T	U	V	W	X	Y	Z	AA	AB
1	movie_imdb_link	num_user_for_reviews	language	country	content_rating	budget	title_year	actor_2_facebook_likes	imdb_score	aspect_ratio	movie_faceb
3637	http://www.imdb.com/title/tt2796678/?ref_=fn_tt_tt_1	3	English	USA		3300000	2014	521	4.5	16	
4042	http://www.imdb.com/title/tt1414361/?ref_=fn_tt_tt_1	28	English	USA	R	1000000	2009	157	2.8	16	
4060	http://www.imdb.com/title/tt1852001/?ref_=fn_tt_tt_1	2	French	France		900000	2011	0	6.7	16	
4109	http://www.imdb.com/title/tt1976608/?ref_=fn_tt_tt_1	5	English	USA	Unrated	500000	2011	169	6.8	16	
4132	http://www.imdb.com/title/tt2275671/?ref_=fn_tt_tt_1	1	Spanish	Dominican Republic	R	500000	2012	15	6.9	16	
4134	http://www.imdb.com/title/tt3846442/?ref_=fn_tt_tt_1	1	English	USA		500000	2016	169	5.2	16	
4241	http://www.imdb.com/title/tt1942884/?ref_=fn_tt_tt_1	23	English	Canada	Not Rated	100000	2012	0	7.7	16	
4259	http://www.imdb.com/title/tt1798291/?ref_=fn_tt_tt_1	7	English	USA	R	42000	2013	571	6.6	16	
4279	http://www.imdb.com/title/tt2049518/?ref_=fn_tt_tt_1	8	English	USA		200000	2012	98	5.4	16	

As we can see the aspect ratio is 16 it is deviation from other values as the fall in unit digit s and decimal points so we convert them into 1.6

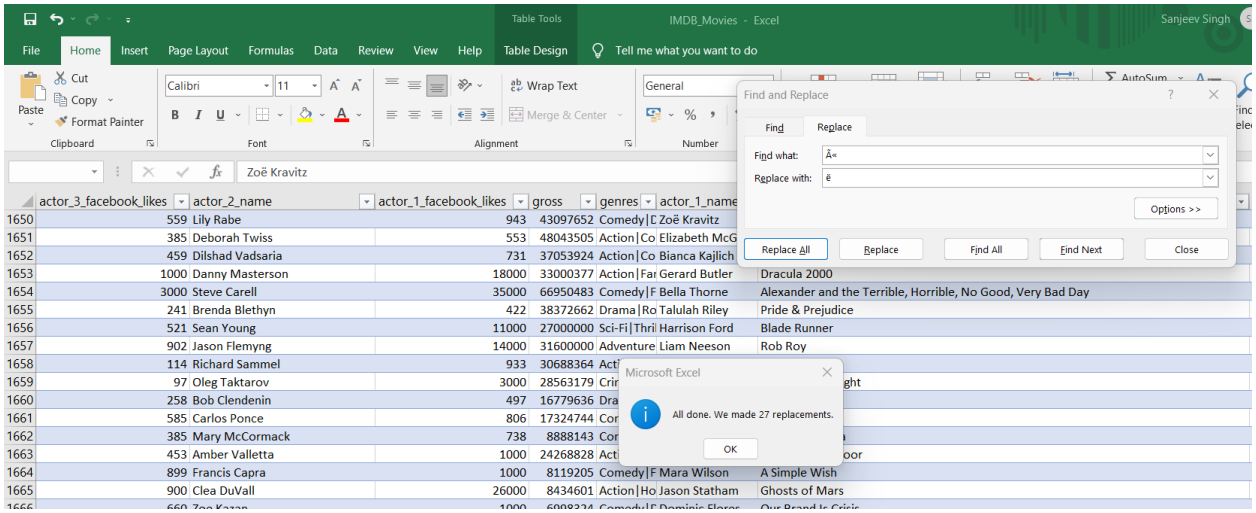


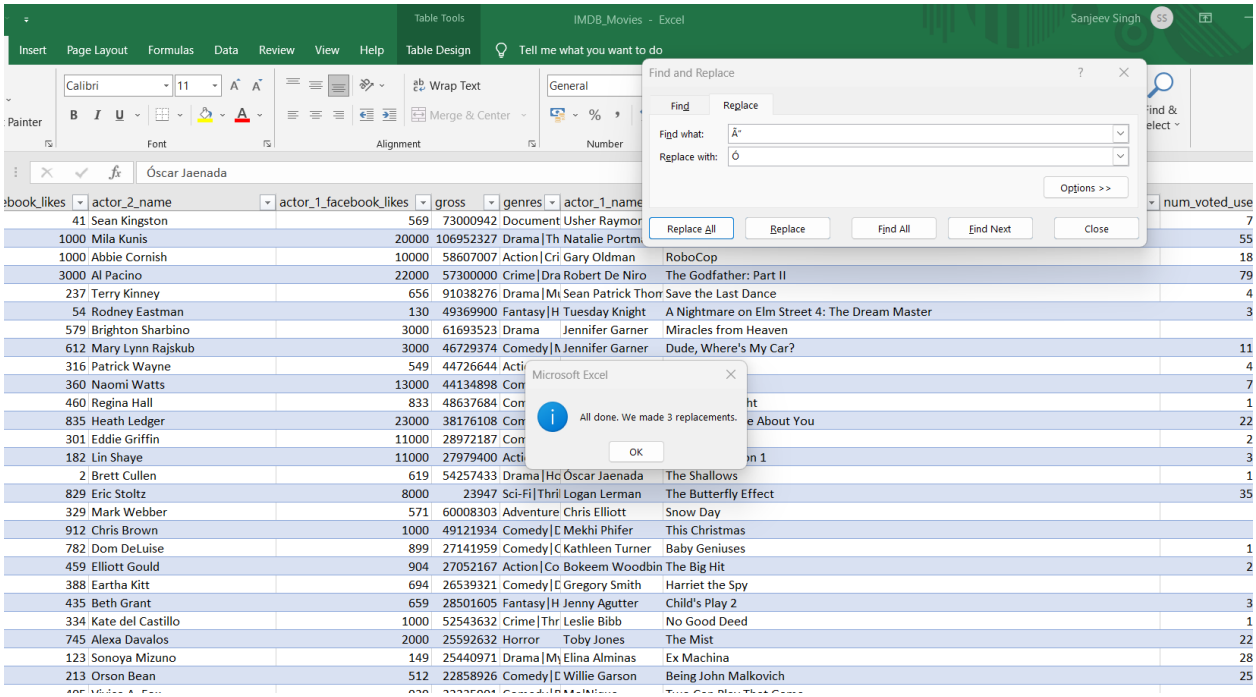
	movie_imdb_link	num_user_for_reviews	language	country	content_rating	budget	title_year	actor_2_facebook_likes	imdb_score	aspect_ratio	movie_facebo
3637	http://www.imdb.com/title/tt2796678/?ref_=fn_tt_tt_1	3	English	USA		3300000	2014	521	4.5	1.6	
4042	http://www.imdb.com/title/tt1414361/?ref_=fn_tt_tt_1	28	English	USA	R	1000000	2009	157	2.8	1.6	
4060	http://www.imdb.com/title/tt1852001/?ref_=fn_tt_tt_1	2	French	France		900000	2011	0	6.7	1.6	
4109	http://www.imdb.com/title/tt1976608/?ref_=fn_tt_tt_1	5	English	USA	Unrated	500000	2011	169	6.8	1.6	
4132	http://www.imdb.com/title/tt2275671/?ref_=fn_tt_tt_1	1	Spanish	Dominican Republic	R	500000	2012	15	6.9	1.6	
4134	http://www.imdb.com/title/tt3846442/?ref_=fn_tt_tt_1	1	English	USA		500000	2016	169	5.2	1.6	
4241	http://www.imdb.com/title/tt1942884/?ref_=fn_tt_tt_1	23	English	Canada	Not Rated	100000	2012	0	7.7	1.6	
4259	http://www.imdb.com/title/tt1798291/?ref_=fn_tt_tt_1	7	English	USA	R	42000	2013	571	6.6	1.6	
4279	http://www.imdb.com/title/tt2049518/?ref_=fn_tt_tt_1	8	English	USA		200000	2012	98	5.4	1.6	

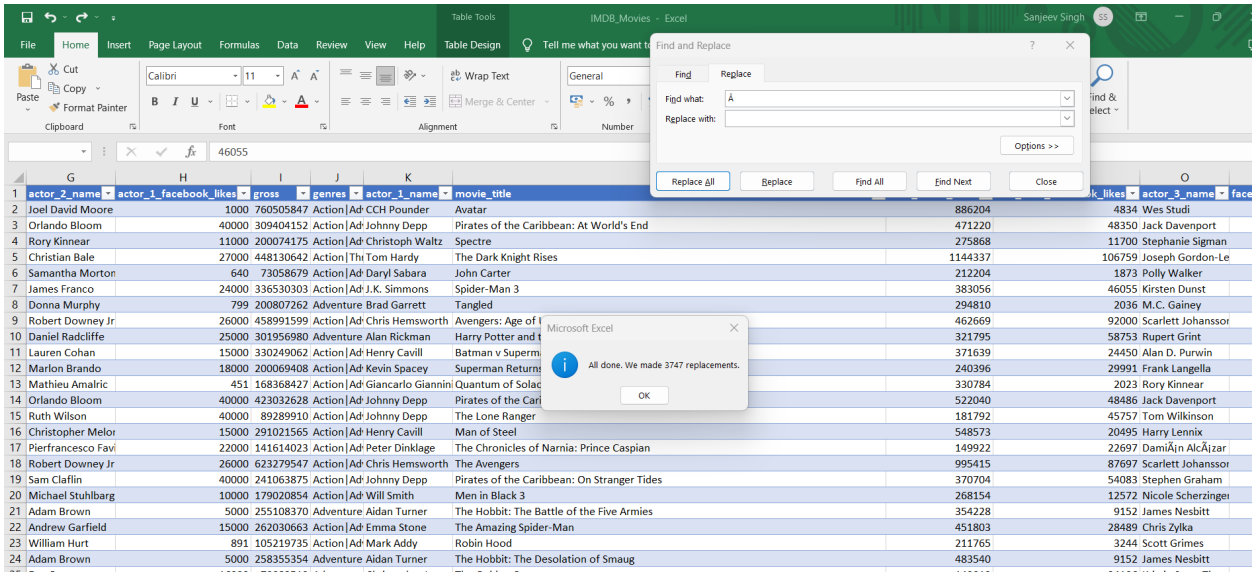
And other replacements as well

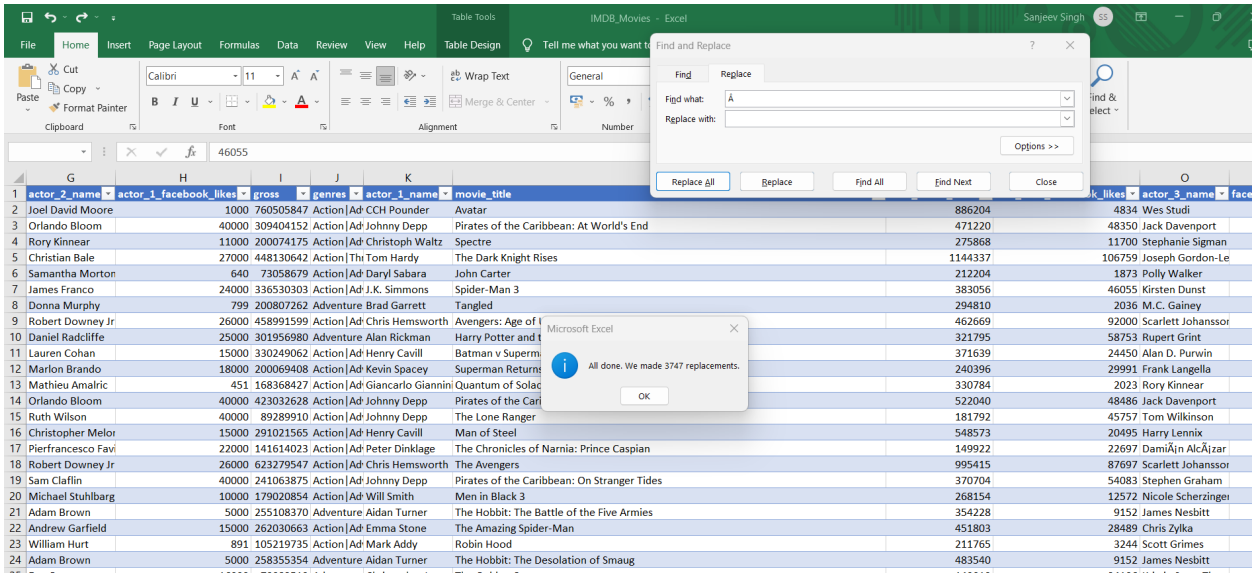


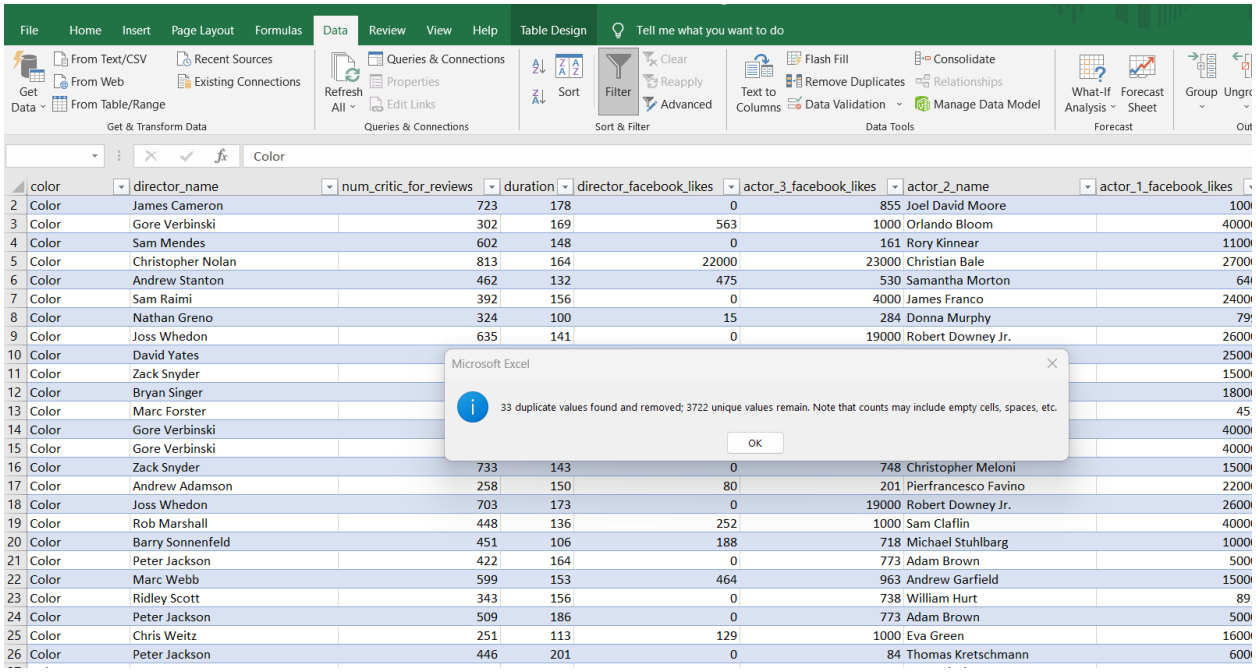
	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
142	Color	Martin Koolhoven	111	103	12	19
143	Color	Prachya Pinkaew	112	111	64	380
144	Color	Gurinder Chadha	161	112	98	397
145	Color	John Sayles	68	141	407	501
146	Color	Preston A. Whitmc	25	95	17	255
147	Color	Jaume BalaguerÃ³	222	85	57	6
148	Color	George Roy Hill	119	129	131	417
149	Color	Hugh Hudson	90	125	46	249
150	Color	Darren Grant	62	116	0	
151	Color	Scott Hicks	71	105	96	
152	Color	Joseph Gordon-Le	364	90	23000	
153	Color	Terry Zwigoff	122	111	72	
154	Color	Richard Eyre	100	91	14	
155	Color	Christophe Barrati	112	97	25	
156	Color	Émile Gaudreault	67	92	9	71
157	Color	James Cox	91	104	10	93
158	Color	Spike Lee	103	120	0	388
159	Color	James Toback	27	99	56	781
160	Color	Aki KaurismÃ¤ki	205	93	592	36
161	Color	Hitoshi Matsumot	72	99	17	16
162	Color	George Ratliff	48	96	3	38
163	Color	David Wain	65	96	136	139
164	Color	Morten Tyldum	284	100	77	27
165	Color	Michael McGowar	47	98	12	149


IMDB_Movies - Excel
Sanjeev Singh
Find and Replace
Find what: Â«
Replace with: ë
Options >>
Replace All
Replace
Find All
Find Next
Close
Microsoft Excel
All done. We made 27 replacements.
OK


IMDB_Movies - Excel
Sanjeev Singh
Find and Replace
Find what: Ã¯
Replace with: Ó
Options >>
Replace All
Replace
Find All
Find Next
Close
Microsoft Excel
All done. We made 3 replacements.
OK


IMDB_Movies - Excel
Sanjeev Singh
Find and Replace
Find what: Â
Replace with:
Options >>
Replace All
Replace
Find All
Find Next
Close
Microsoft Excel
All done. We made 3747 replacements.
OK


IMDB_Movies - Excel
Sanjeev Singh
Find and Replace
Find what: Â
Replace with:
Options >>
Replace All
Replace
Find All
Find Next
Close
Microsoft Excel
All done. We made 3747 replacements.
OK

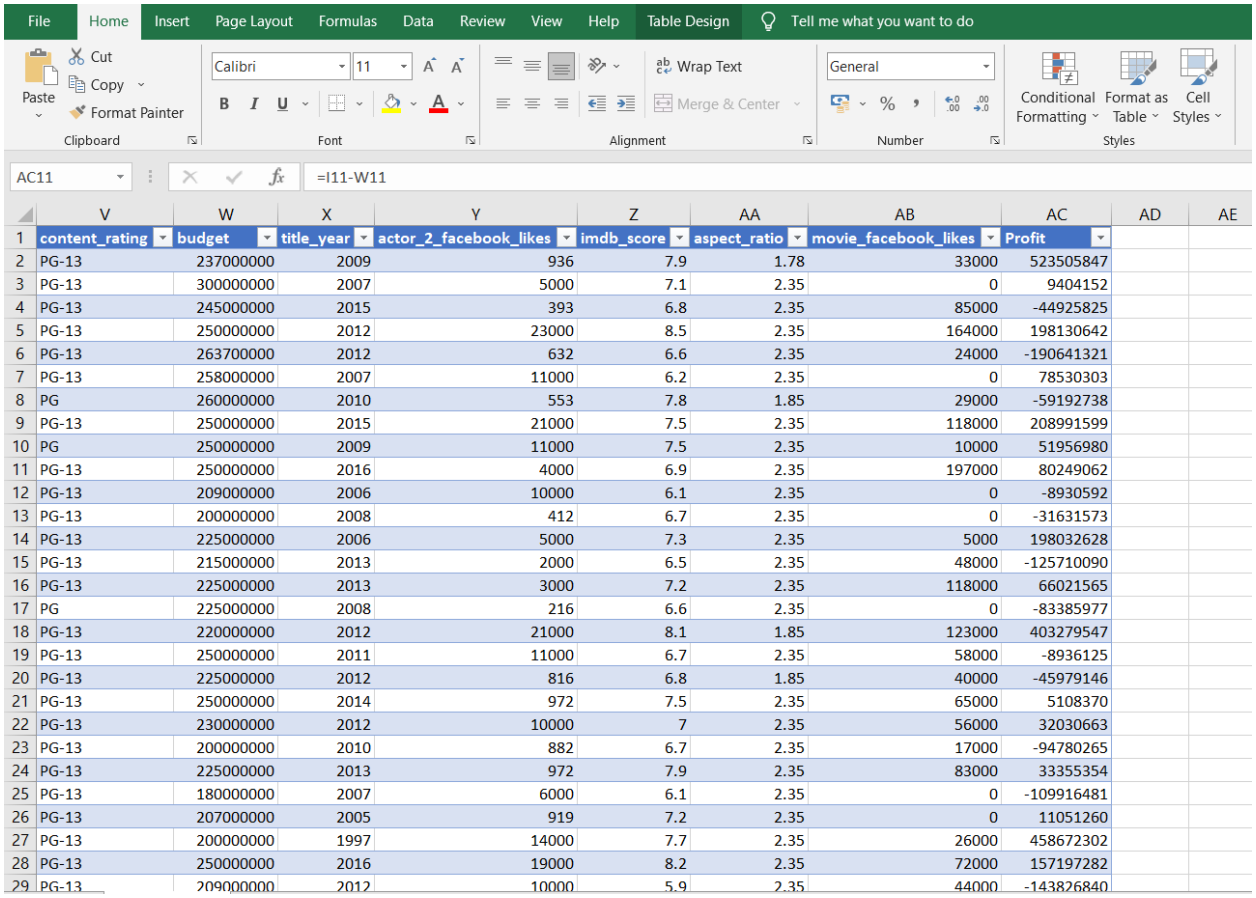


	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes
2	Color	James Cameron	723	178	0	855	Joel David Moore	100
3	Color	Gore Verbinski	302	169	563	1000	Orlando Bloom	4000
4	Color	Sam Mendes	602	148	0	161	Rory Kinnear	1100
5	Color	Christopher Nolan	813	164	22000	23000	Christian Bale	2700
6	Color	Andrew Stanton	462	132	475	530	Samantha Morton	64
7	Color	Sam Raimi	392	156	0	4000	James Franco	2400
8	Color	Nathan Greno	324	100	15	284	Donna Murphy	79
9	Color	Joss Whedon	635	141	0	19000	Robert Downey Jr.	2600
10	Color	David Yates						2500
11	Color	Zack Snyder						1500
12	Color	Bryan Singer						1800
13	Color	Marc Forster						45
14	Color	Gore Verbinski						4000
15	Color	Gore Verbinski						4000
16	Color	Zack Snyder	733	143	0	748	Christopher Meloni	1500
17	Color	Andrew Adamson	258	150	80	201	Pierfrancesco Favino	2200
18	Color	Joss Whedon	703	173	0	19000	Robert Downey Jr.	2600
19	Color	Rob Marshall	448	136	252	1000	Sam Claflin	4000
20	Color	Barry Sonnenfeld	451	106	188	718	Michael Stuhlbarg	1000
21	Color	Peter Jackson	422	164	0	773	Adam Brown	500
22	Color	Marc Webb	599	153	464	963	Andrew Garfield	1500
23	Color	Ridley Scott	343	156	0	738	William Hurt	89
24	Color	Peter Jackson	509	186	0	773	Adam Brown	500
25	Color	Chris Weitz	251	113	129	1000	Eva Green	1600
26	Color	Peter Jackson	446	201	0	84	Thomas Kretschmann	600

And removing duplicates (other various changes that are minute to be documented but most;y spelling deviations)

This helps us clear the data

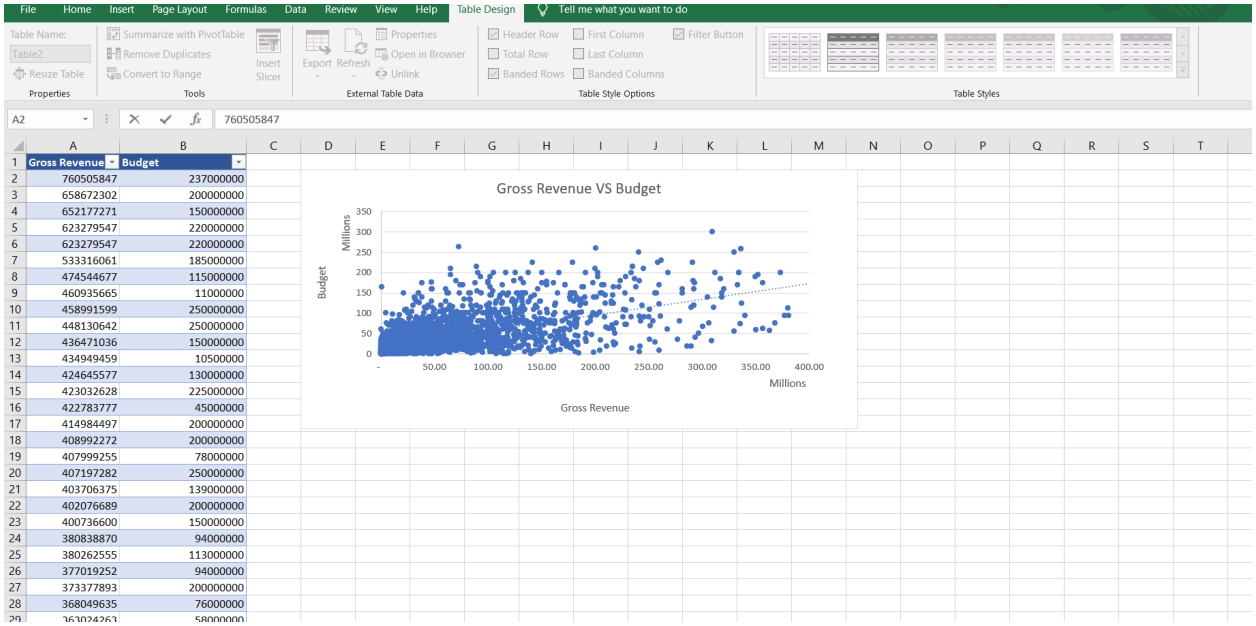
2. Movies with highest profit: Create a new column called profit which contains the difference of the two columns: gross and budget. Sort the column using the profit column as reference. Plot profit (y-axis) vs budget (x- axis) and observe the outliers using the appropriate chart type.
Your task: Find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



	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	content_rating	budget	title_year	actor_2_facebook_likes	imdb_score	aspect_ratio	movie_facebook_likes	Profit		
2	PG-13	237000000	2009	936	7.9	1.78	33000	523505847		
3	PG-13	300000000	2007	5000	7.1	2.35	0	9404152		
4	PG-13	245000000	2015	393	6.8	2.35	85000	-44925825		
5	PG-13	250000000	2012	23000	8.5	2.35	164000	198130642		
6	PG-13	263700000	2012	632	6.6	2.35	24000	-190641321		
7	PG-13	258000000	2007	11000	6.2	2.35	0	78530303		
8	PG	260000000	2010	553	7.8	1.85	29000	-59192738		
9	PG-13	250000000	2015	21000	7.5	2.35	118000	208991599		
10	PG	250000000	2009	11000	7.5	2.35	10000	51956980		
11	PG-13	250000000	2016	4000	6.9	2.35	197000	80249062		
12	PG-13	209000000	2006	10000	6.1	2.35	0	-8930592		
13	PG-13	200000000	2008	412	6.7	2.35	0	-31631573		
14	PG-13	225000000	2006	5000	7.3	2.35	5000	198032628		
15	PG-13	215000000	2013	2000	6.5	2.35	48000	-125710090		
16	PG-13	225000000	2013	3000	7.2	2.35	118000	66021565		
17	PG	225000000	2008	216	6.6	2.35	0	-83385977		
18	PG-13	220000000	2012	21000	8.1	1.85	123000	403279547		
19	PG-13	250000000	2011	11000	6.7	2.35	58000	-8936125		
20	PG-13	225000000	2012	816	6.8	1.85	40000	-45979146		
21	PG-13	250000000	2014	972	7.5	2.35	65000	5108370		
22	PG-13	230000000	2012	10000	7	2.35	56000	32030663		
23	PG-13	200000000	2010	882	6.7	2.35	17000	-94780265		
24	PG-13	225000000	2013	972	7.9	2.35	83000	33355354		
25	PG-13	180000000	2007	6000	6.1	2.35	0	-109916481		
26	PG-13	207000000	2005	919	7.2	2.35	0	11051260		
27	PG-13	200000000	1997	14000	7.7	2.35	26000	458672302		
28	PG-13	250000000	2016	19000	8.2	2.35	72000	157197282		
29	PG-13	209000000	2012	10000	5.9	2.35	44000	-143826840		

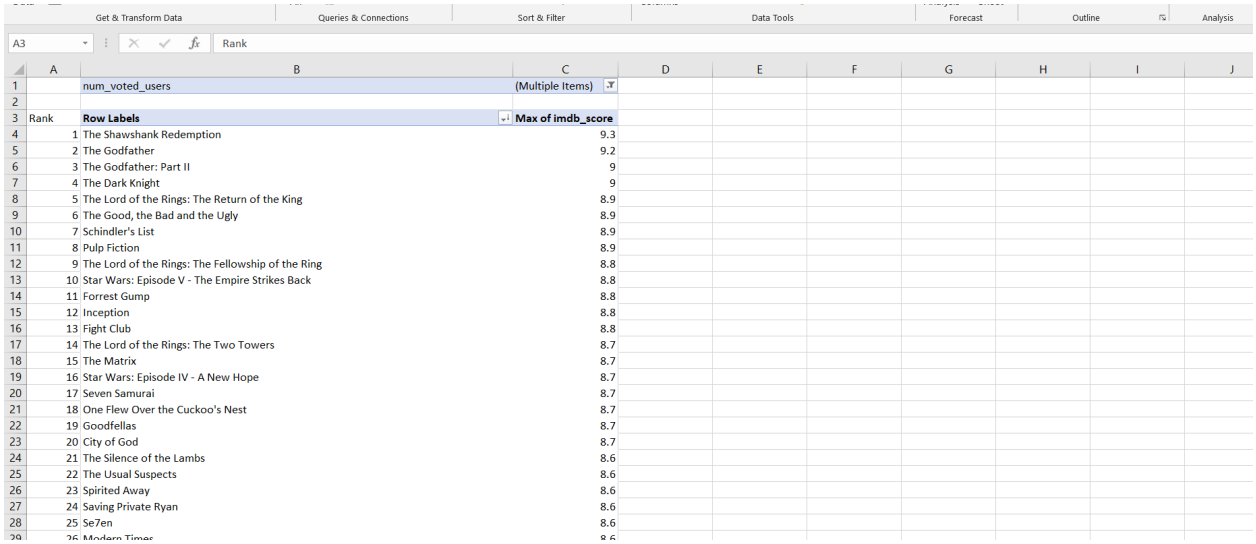
Insight: Convert the unit of the budget and gross columns from $ to million $ by dividing the amount by 1000000

Based on table of profit we create x and y axis graph



Gross Revenue	Budget
760505847	237000000
658672302	200000000
652177271	150000000
623279547	220000000
623279547	220000000
533316061	185000000
474544677	115000000
460935665	11000000
458991599	250000000
448130642	250000000
436471036	150000000
434949459	10500000
424645577	130000000
423032628	225000000
422783777	45000000
414984497	200000000
408992272	200000000
407999255	78000000
407197282	250000000
403706375	139000000
402076689	200000000
400736600	150000000
380838870	94000000
380262555	113000000
377019252	94000000
373377893	200000000
368049635	76000000

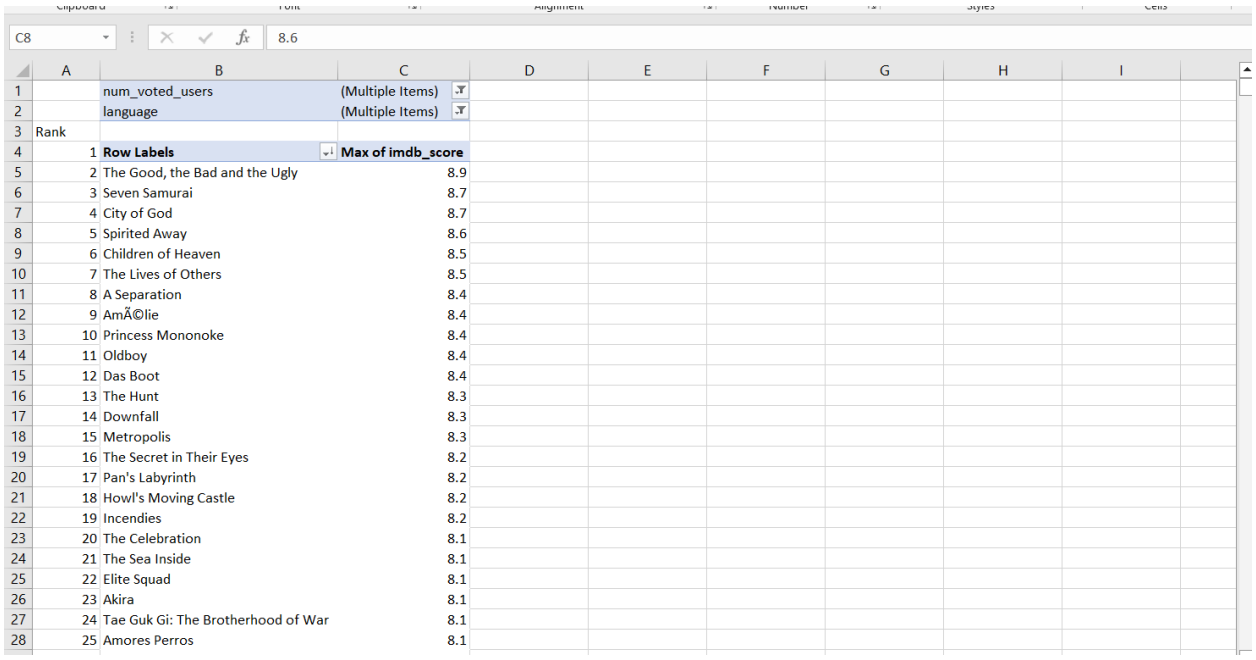
3. Top 250: Create a new column IMDb_Top_250 and store the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb_score). Also make sure that for all of these movies, the num_voted_users is greater than 25,000. Also add a Rank column containing the values 1 to 250 indicating the ranks of the corresponding films.



Rank	Row Labels	Max of imdb_score
1	The Shawshank Redemption	9.3
2	The Godfather	9.2
3	The Godfather: Part II	9
4	The Dark Knight	9
5	The Lord of the Rings: The Return of the King	8.9
6	The Good, the Bad and the Ugly	8.9
7	Schindler's List	8.9
8	Pulp Fiction	8.9
9	The Lord of the Rings: The Fellowship of the Ring	8.8
10	Star Wars: Episode V - The Empire Strikes Back	8.8
11	Forrest Gump	8.8
12	Inception	8.8
13	Fight Club	8.8
14	The Lord of the Rings: The Two Towers	8.7
15	The Matrix	8.7
16	Star Wars: Episode IV - A New Hope	8.7
17	Seven Samurai	8.7
18	One Flew Over the Cuckoo's Nest	8.7
19	Goodfellas	8.7
20	City of God	8.7
21	The Silence of the Lambs	8.6
22	The Usual Suspects	8.6
23	Spirited Away	8.6
24	Saving Private Ryan	8.6
25	Se7en	8.6

Using pivot table and get imdb score for to get 25,000 above we apply filers using pivot table `

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.



Rank	Row Labels	Max of imdb_score
2	The Good, the Bad and the Ugly	8.9
3	Seven Samurai	8.7
4	City of God	8.7
5	Spirited Away	8.6
6	Children of Heaven	8.5
7	The Lives of Others	8.5
8	A Separation	8.4
9	AmÃ©lie	8.4
10	Princess Mononoke	8.4
11	Oldboy	8.4
12	Das Boot	8.4
13	The Hunt	8.3
14	Downfall	8.3
15	Metropolis	8.3
16	The Secret in Their Eyes	8.2
17	Pan's Labyrinth	8.2
18	Howl's Moving Castle	8.2
19	Incendies	8.2
20	The Celebration	8.1
21	The Sea Inside	8.1
22	Elite Squad	8.1
23	Akira	8.1
24	Tae Guk Gi: The Brotherhood of War	8.1
25	Amores Perros	8.1

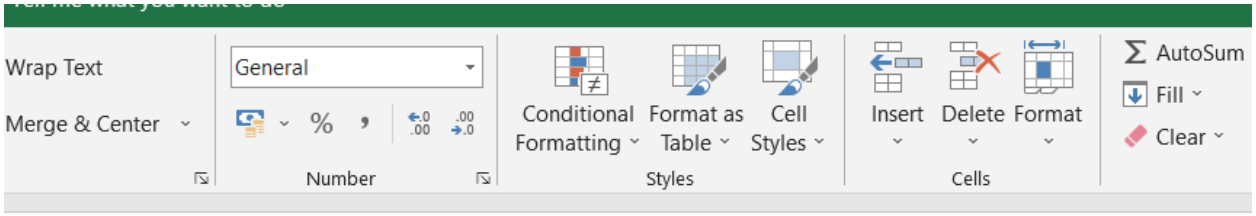
By adding language as filter and including all languages except english

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.

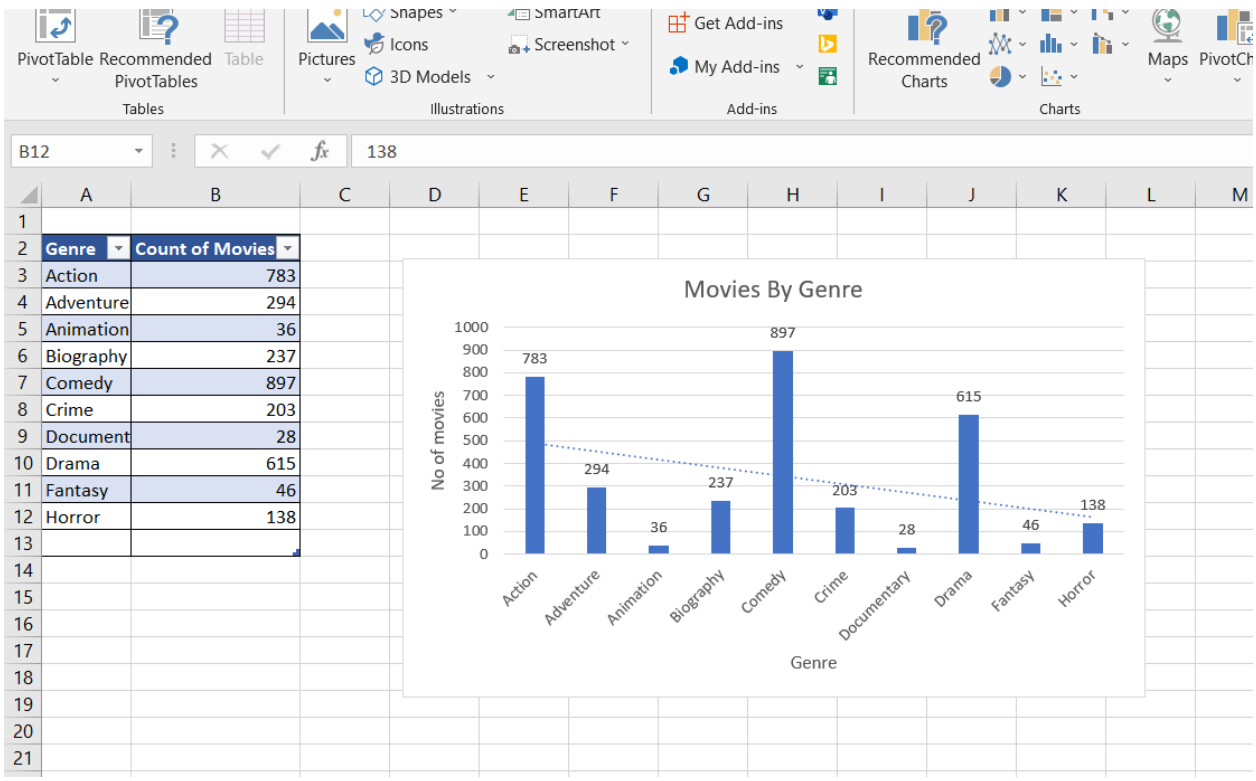
Row Labels	Average of imdb_score
Akira Kurosawa	8.7
Tony Kaye	8.6
Charles Chaplin	8.6
Ron Fricke	8.5
Majid Majidi	8.5
Damien Chazelle	8.5
Alfred Hitchcock	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
Richard Marquand	8.4
Asghar Farhadi	8.4
Lenny Abrahamson	8.3
Lee Unkrich	8.3
Fritz Lang	8.3
Billy Wilder	8.3
Pete Docter	8.233333333
Hayao Miyazaki	8.225

Top 10 director based on imdb score and then sorting them alphabetically



Rank	Director	imdb
1	Akira Kurosawa	8.7
2	Charles Chaplin	8.6
3	Tony Kaye	8.6
4	Alfred Hitchcock	8.5
5	Damien Chazelle	8.5
6	Majid Majidi	8.5
7	Ron Fricke	8.5
8	Sergio Leone	8.433333
9	Christopher Nolan	8.425
10	Asghar Farhadi	8.4

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



Genre	Count of Movies
Action	783
Adventure	294
Animation	36
Biography	237
Comedy	897
Crime	203
Document	28
Drama	615
Fantasy	46
Horror	138

insight:I have split the values from the genres column and extracted them into 3 columns

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.

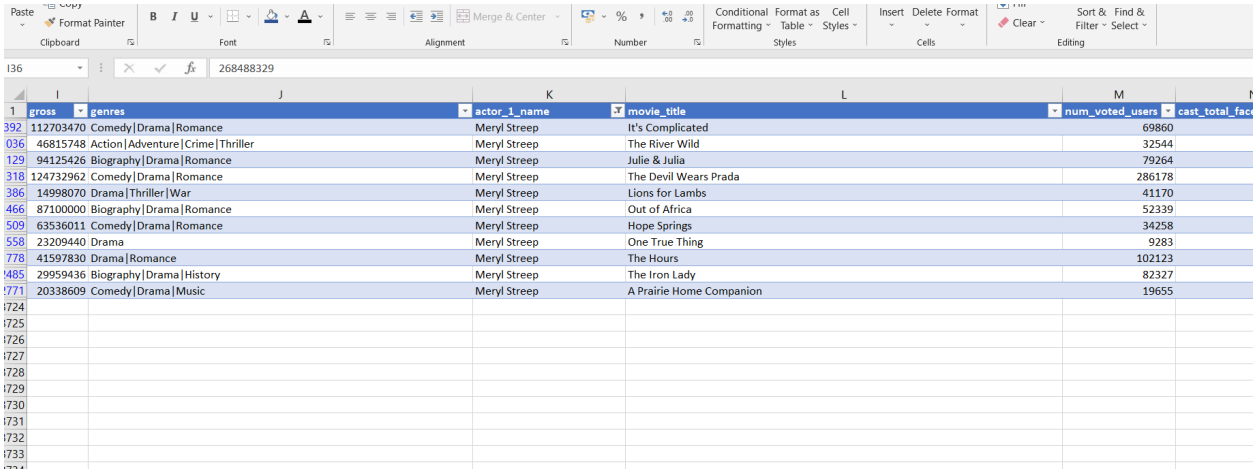
Group the combined column using the actor_1_name column.

Find the mean of the num_critic_for_reviews and num_users_for_review and identify the actors which have the highest mean.

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	L	M	N
1	gross	genres	actor_1_name	movie_title	num_voted_users	cast_total_fac
392	112703470	Comedy|Drama|Romance	Meryl Streep	It's Complicated	69860	
036	46815748	Action|Adventure|Crime|Thriller	Meryl Streep	The River Wild	32544	
129	94125426	Biography|Drama|Romance	Meryl Streep	Julie & Julia	79264	
318	124732962	Comedy|Drama|Romance	Meryl Streep	The Devil Wears Prada	286178	
386	14998070	Drama|Thriller|War	Meryl Streep	Lions for Lambs	41170	
466	87100000	Biography|Drama|Romance	Meryl Streep	Out of Africa	52339	
509	63536011	Comedy|Drama|Romance	Meryl Streep	Hope Springs	34258	
558	23209440	Drama	Meryl Streep	One True Thing	9283	
778	41597830	Drama|Romance	Meryl Streep	The Hours	102123	
485	29959436	Biography|Drama|History	Meryl Streep	The Iron Lady	82327	
771	20338609	Comedy|Drama|Music	Meryl Streep	A Prairie Home Companion	19655	
724						
725						
726						
727						
728						
729						
730						
731						
732						
733						

Meryl streep

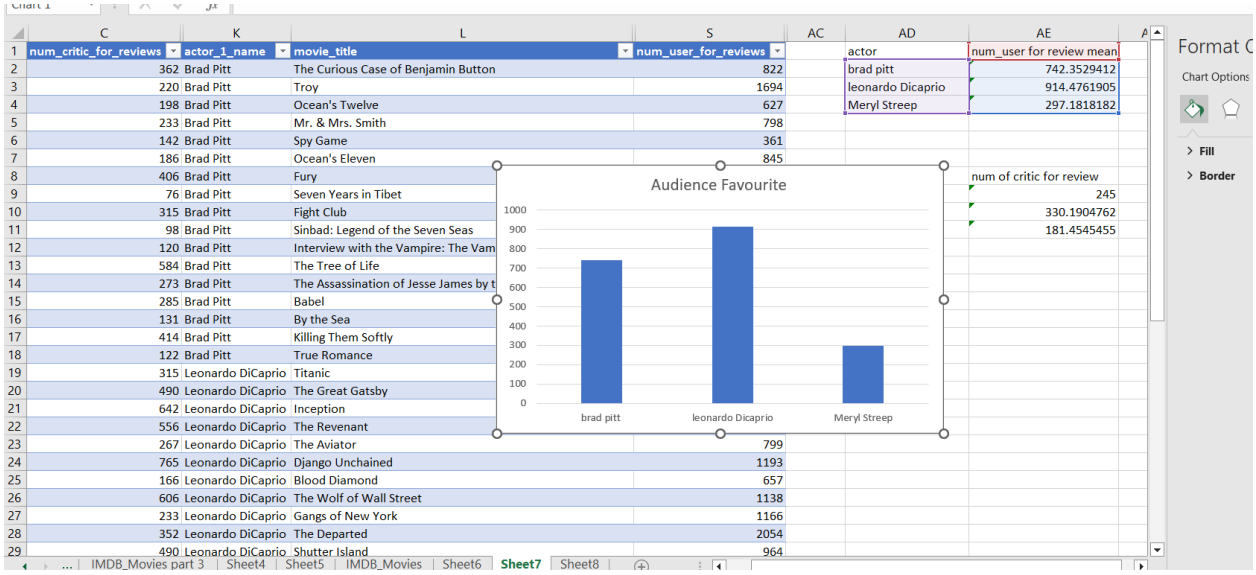
I	J	K	L	M	N
gross	genres	actor_1_name	movie_title	num_voted_users	cast_total_facebo
658672302	Drama|Romance	Leonardo DiCaprio	Titanic	793059	
144812796	Drama|Romance	Leonardo DiCaprio	The Great Gatsby	362912	
292568851	Action|Adventure|Sci-Fi|Thriller	Leonardo DiCaprio	Inception	1468200	
183635922	Adventure|Drama|Thriller|Western	Leonardo DiCaprio	The Revenant	406020	
102608827	Biography|Drama	Leonardo DiCaprio	The Aviator	264318	
162804648	Drama|Western	Leonardo DiCaprio	Django Unchained	955174	
57366262	Adventure|Drama|Thriller	Leonardo DiCaprio	Blood Diamond	400292	
116866727	Biography|Comedy|Crime|Drama	Leonardo DiCaprio	The Wolf of Wall Street	780588	
77679638	Crime|Drama	Leonardo DiCaprio	Gangs of New York	314033	
132373442	Crime|Drama|Thriller	Leonardo DiCaprio	The Departed	873649	
127968405	Mystery|Thriller	Leonardo DiCaprio	Shutter Island	786092	
39380442	Action|Drama|Thriller	Leonardo DiCaprio	Body of Lies	174248	
164435221	Biography|Crime|Drama	Leonardo DiCaprio	Catch Me If You Can	525801	
39778599	Adventure|Drama|Thriller	Leonardo DiCaprio	The Beach	176169	
22877808	Drama|Romance	Leonardo DiCaprio	Revolutionary Road	152591	
56876365	Action|Adventure	Leonardo DiCaprio	The Man in the Iron Mask	125219	
37304950	Biography|Crime|Drama	Leonardo DiCaprio	J. Edgar	102728	
18636537	Action|Thriller|Western	Leonardo DiCaprio	The Quick and the Dead	69197	
12782508	Drama	Leonardo DiCaprio	Marvin's Room	20163	
46338728	Drama|Romance	Leonardo DiCaprio	Romeo + Juliet	167750	
144812796	Drama|Romance	Leonardo DiCaprio	The Great Gatsby	362933	

Leonardo Dicaprio

I	J	K	L	M	N
gross	genres	actor_1_name	movie_title	num_voted_users	cast_total_faceb
127490802	Drama|Fantasy|Romance	Brad Pitt	The Curious Case of Benjamin Button	459346	
133228348	Adventure	Brad Pitt	Troy	381672	
125531634	Crime|Thriller	Brad Pitt	Ocean's Twelve	284852	
186336103	Action|Comedy|Crime|Romance|Thriller	Brad Pitt	Mr. & Mrs. Smith	348861	
26871	Action|Crime|Thriller	Brad Pitt	Spy Game	121259	
183405771	Crime|Thriller	Brad Pitt	Ocean's Eleven	402645	
85707116	Action|Drama|War	Brad Pitt	Fury	303185	
37901509	Adventure|Biography|Drama|History|War	Brad Pitt	Seven Years in Tibet	96385	
37023395	Drama	Brad Pitt	Fight Club	1347461	
26288320	Adventure|Animation|Comedy|Drama|Family|Fantasy|Romance	Brad Pitt	Sinbad: Legend of the Seven Seas	36144	
105264608	Drama|Fantasy|Horror	Brad Pitt	Interview with the Vampire: The Vampire Chronicles	239752	
13303319	Drama|Fantasy	Brad Pitt	The Tree of Life	136367	
3904982	Biography|Crime|Drama|History|Western	Brad Pitt	The Assassination of Jesse James by the Coward Robert Ford	136104	
34300771	Drama	Brad Pitt	Babel	243799	
531009	Drama|Romance	Brad Pitt	By the Sea	7976	
14938570	Crime|Thriller	Brad Pitt	Killing Them Softly	111625	
12281500	Action|Crime|Drama|Romance|Thriller	Brad Pitt	True Romance	163492	

Brad pitt

	Meryl steep	leonardo dicaprio	Brad Pitt	Combined	E	F	G	H	I
4	Julie & Julia	Inception	Ocean's Twelve	Julie & Julia					
5	The Devil Wears Prada	The Revenant	Mr. & Mrs. Smith	The Devil Wears Prada					
6	Lions for Lambs	The Aviator	Spy Game	Lions for Lambs					
7	Out of Africa	Django Unchained	Ocean's Eleven	Out of Africa					
8	Hope Springs	Blood Diamond	Fury	Hope Springs					
9	One True Thing	The Wolf of Wall Street	Seven Years in Tibet	One True Thing					
10	The Hours	Gangs of New York	Fight Club	The Hours					
11	The Iron Lady	The Departed	Sinbad: Legend of the Seven Seas	The Iron Lady					
12	A Prairie Home Companion	Shutter Island	Interview with the Vampire: The Vampire Chronicles	A Prairie Home Companion					
13		Body of Lies	The Tree of Life	Titanic					
14		Catch Me If You Can	The Assassination of Jesse James by the Coward Robert Ford	The Great Gatsby					
15		The Beach	Babel	Inception					
16		Revolutionary Road	By the Sea	The Revenant					
17		The Man in the Iron Mask	Killing Them Softly	The Aviator					
18		J. Edgar	True Romance	Django Unchained					
19		The Quick and the Dead		Blood Diamond					
20		Marvin's Room		The Wolf of Wall Street					
21		Romeo + Juliet		Gangs of New York					
22		The Great Gatsby		The Departed					
23				Shutter Island					
24				Body of Lies					
25				Catch Me If You Can					
26				The Beach					
27				Revolutionary Road					
28				The Man in the Iron Mask					
29				J. Edgar					
30				The Quick and the Dead					
31				Marvin's Room					

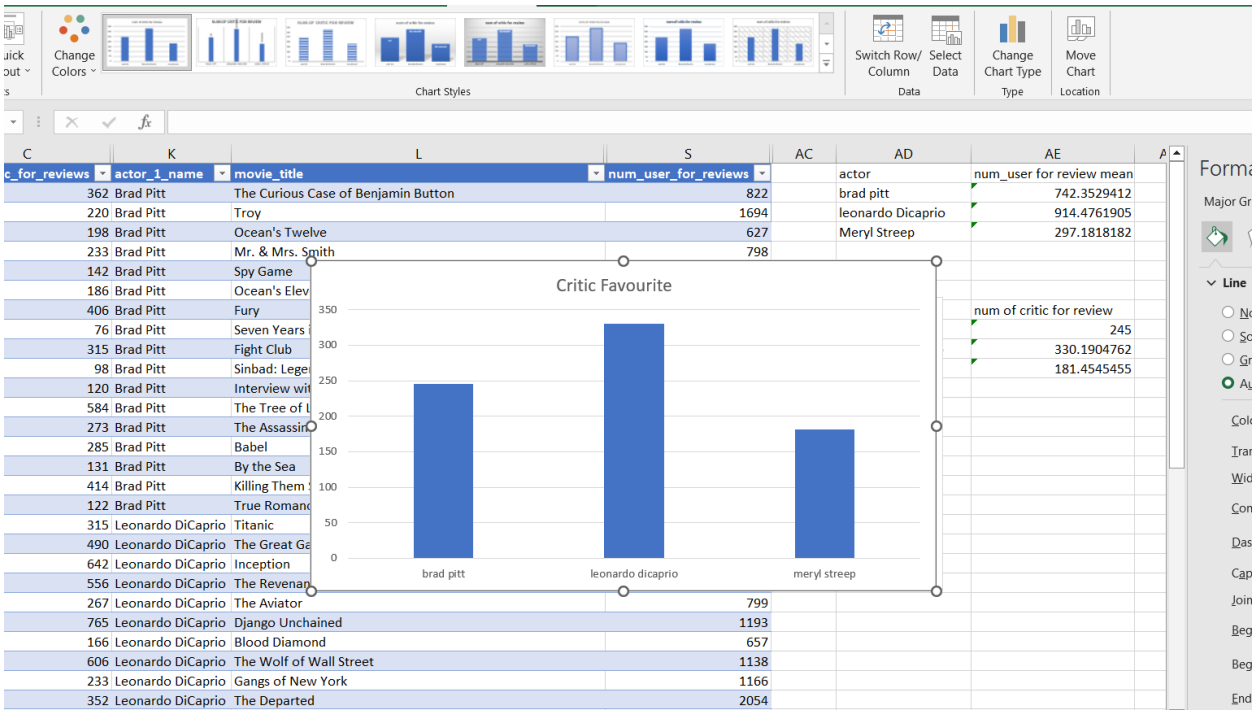


num_critic_for_reviews	actor_1_name	movie_title	num_user_for_reviews
362	Brad Pitt	The Curious Case of Benjamin Button	822
220	Brad Pitt	Troy	1694
198	Brad Pitt	Ocean's Twelve	627
233	Brad Pitt	Mr. & Mrs. Smith	798
142	Brad Pitt	Spy Game	361
186	Brad Pitt	Ocean's Eleven	845
406	Brad Pitt	Fury	
76	Brad Pitt	Seven Years in Tibet	
315	Brad Pitt	Fight Club	
98	Brad Pitt	Sinbad: Legend of the Seven Seas	
120	Brad Pitt	Interview with the Vampire: The Vam	
584	Brad Pitt	The Tree of Life	
273	Brad Pitt	The Assassination of Jesse James by t	
285	Brad Pitt	Babel	
131	Brad Pitt	By the Sea	
414	Brad Pitt	Killing Them Softly	
122	Brad Pitt	True Romance	
315	Leonardo DiCaprio	Titanic	
490	Leonardo DiCaprio	The Great Gatsby	
642	Leonardo DiCaprio	Inception	
556	Leonardo DiCaprio	The Revenant	
267	Leonardo DiCaprio	The Aviator	799
765	Leonardo DiCaprio	Django Unchained	1193
166	Leonardo DiCaprio	Blood Diamond	657
606	Leonardo DiCaprio	The Wolf of Wall Street	1138
233	Leonardo DiCaprio	Gangs of New York	1166
352	Leonardo DiCaprio	The Departed	2054
490	Leonardo DiCaprio	Shutter Island	964

actor	num_user for review mean
brad pitt	742.3529412
leonardo Dicaprio	914.4761905
Meryl Streep	297.1818182

num of critic for review
245
330.1904762
181.4545455

Audience fav



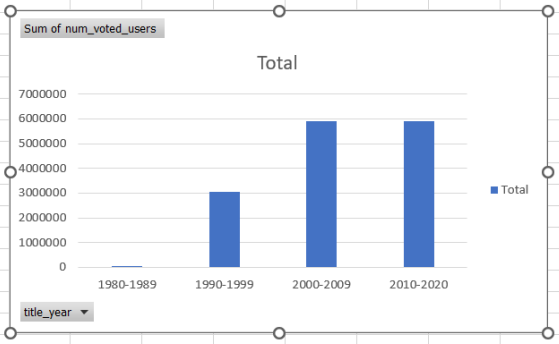
num_critic_for_reviews	actor_1_name	movie_title	num_user_for_reviews
362	Brad Pitt	The Curious Case of Benjamin Button	822
220	Brad Pitt	Troy	1694
198	Brad Pitt	Ocean's Twelve	627
233	Brad Pitt	Mr. & Mrs. Smith	798
142	Brad Pitt	Spy Game	
186	Brad Pitt	Ocean's Elev	
406	Brad Pitt	Fury	
76	Brad Pitt	Seven Years	
315	Brad Pitt	Fight Club	
98	Brad Pitt	Sinbad: Lege	
120	Brad Pitt	Interview wit	
584	Brad Pitt	The Tree of L	
273	Brad Pitt	The Assassin	
285	Brad Pitt	Babel	
131	Brad Pitt	By the Sea	
414	Brad Pitt	Killing Them	
122	Brad Pitt	True Romanc	
315	Leonardo DiCaprio	Titanic	
490	Leonardo DiCaprio	The Great Ga	
642	Leonardo DiCaprio	Inception	
556	Leonardo DiCaprio	The Revenan	
267	Leonardo DiCaprio	The Aviator	799
765	Leonardo DiCaprio	Django Unchained	1193
166	Leonardo DiCaprio	Blood Diamond	657
606	Leonardo DiCaprio	The Wolf of Wall Street	1138
233	Leonardo DiCaprio	Gangs of New York	1166
352	Leonardo DiCaprio	The Departed	2054

actor	num_user for review mean
brad pitt	742.3529412
leonardo Dicaprio	914.4761905
Meryl Streep	297.1818182

num of critic for review
245
330.1904762
181.4545455

Critic Fav

Row Labels	Sum of num_voted_users
1980-1989	52339
1990-1999	3064305
2000-2009	5894033
2010-2020	5900385
Grand Total	**14911062**



Insight: By pivot table we extract title year and num of voted users and then by grouping them in decade find the trend