

```
In [12]:
import pandas as pd
import numpy as np
import seaborn as sns

import matplotlib.pyplot as plt
import matplotlib.mlab as mlab
import matplotlib
plt.style.use('ggplot')
from matplotlib.pyplot import figure

%matplotlib inline
matplotlib.rcParams['figure.figsize'] = (12,8)

pd.options.mode.chained_assignment = None

df = pd.read_csv(r"C:\Users\ADMIN\Downloads\movies.csv.zip")

In [5]:
# To Look at the data
df.head()
```

```
Out[5]:
```

	name	rating	genre	year	released	score	votes	director	writer	star	country	budget	
0	The Shining	R	Drama	1980	June 13, 1980 (United States)	8.4	927000.0	Stanley Kubrick	Stephen King	Jack Nicholson	United Kingdom	19000000.0	46996
1	The Blue Lagoon	R	Adventure	1980	July 2, 1980 (United States)	5.8	65000.0	Randal Kleiser	Henry De Vere Stacpoole	Brooke Shields	United States	4500000.0	58853
2	Star Wars: Episode V - The Empire Strikes Back	PG	Action	1980	June 20, 1980 (United States)	8.7	1200000.0	Irvin Kershner	Leigh Brackett	Mark Hamill	United States	18000000.0	538375
3	Airplane!	PG	Comedy	1980	July 2, 1980 (United States)	7.7	221000.0	Jim Abrahams	Jim Abrahams	Robert Hays	United States	3500000.0	83453
4	Caddyshack	R	Comedy	1980	July 25, 1980 (United States)	7.3	108000.0	Harold Ramis	Brian Doyle-Murray	Chevy Chase	United States	6000000.0	39846

```
In [6]:
# To check any missing data

for col in df.columns:
    pct_missing = np.mean(df[col].isnull())
    print('{} - {}'.format(col, pct_missing))
```

```
name - 0.0%
rating - 0.018041731872717789%
genre - 0.0%
year - 0.0%
released - 0.00025082632044861763%
score - 0.0003912363867292645%
votes - 0.0003912363867292645%
director - 0.0%
writer - 0.0003912363867292645%
star - 0.00013841218224388815%
country - 0.0003912363867292645%
budget - 0.2831246739657444%
gross - 0.02464788732394366%
company - 0.002217085738132499%
runtime - 0.0005216484089723526%
```

```
In [8]:
#Data Types for the columns

df.dtypes
```

```
Out[8]:
```

name	object
rating	object
genre	object
year	int64
released	object
score	float64
votes	float64
director	object
writer	object
star	object
country	object
budget	float64
gross	float64
company	object
runtime	float64
dtype:	object

```
In [10]:
# Drop rows with NaN values in 'budget' or 'gross'
df = df.dropna(subset=['budget', 'gross'])

# Convert columns to integers
df['budget'] = df['budget'].astype('int64')
df['gross'] = df['gross'].astype('int64')
```

```
In [20]:
df.head()
```

```
Out[20]:
```

	name	rating	genre	year	released	score	votes	director	writer	star	country	budget	gross	company	n
0	6587	6	6	1980	1705	8.4	927000.0	2589	4014	1047	54	19000000.0	46998772.0		2319
1	5573	6	1	1980	1492	5.8	65000.0	2269	1632	327	55	4500000.0	58853106.0		731
2	5142	4	0	1980	1771	8.7	1200000.0	1111	2567	1745	55	18000000.0	538375067.0		1540
3	286	4	4	1980	1492	7.7	221000.0	1301	2000	2246	55	3500000.0	83453539.0		1812
4	1027	6	4	1980	1543	7.3	108000.0	1054	521	410	55	6000000.0	39846344.0		1777

```
In [30]:
df['yearcorrect'] = df['released'].astype(str).str[-4]

df.head()
```

```
Out[30]:
```

	name	rating	genre	year	released	score	votes	director	writer	star	country	budget	gross	company	n
0	6587	6	6	1980	1705	8.4	927000.0	2589	4014	1047	54	19000000.0	46998772.0		2319
1	5573	6	1	1980	1492	5.8	65000.0	2269	1632	327	55	4500000.0	58853106.0		731
2	5142	4	0	1980	1771	8.7	1200000.0	1111	2567	1745	55	18000000.0	538375067.0		1540
3	286	4	4	1980	1492	7.7	221000.0	1301	2000	2246	55	3500000.0	83453539.0		1812
4	1027	6	4	1980	1543	7.3	108000.0	1054	521	410	55	6000000.0	39846344.0		1777

```
In [22]:
df = df.sort_values(by=['gross'], inplace=False, ascending=False)
```

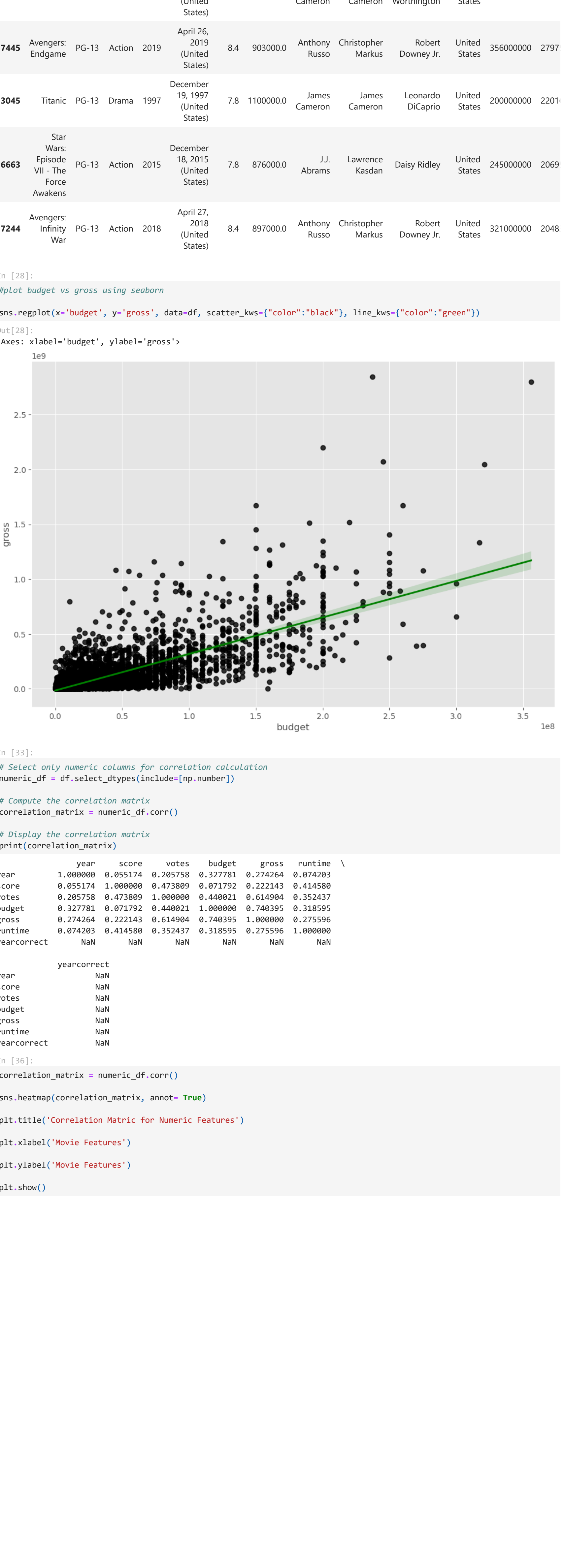
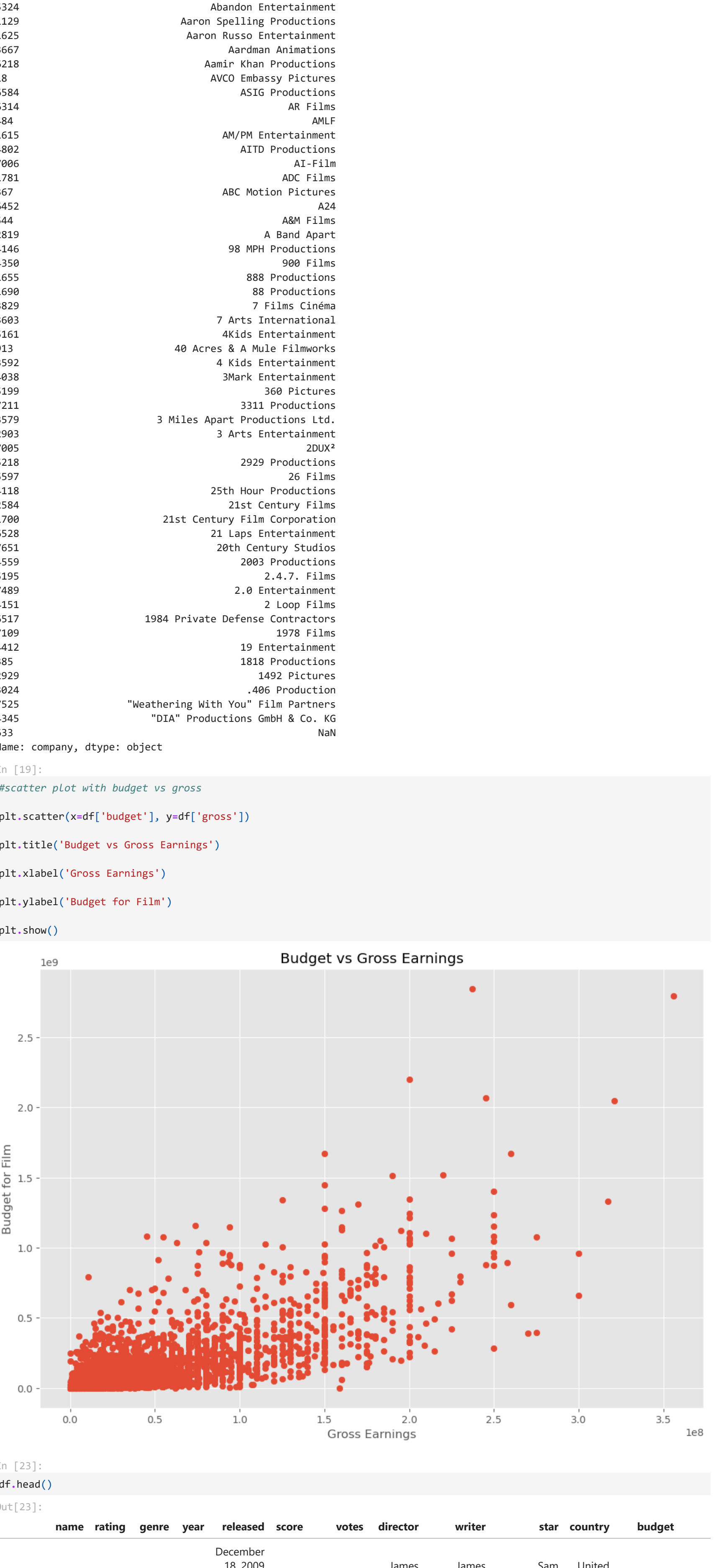
```
In [16]:
pd.set_option('display.max_rows', None)
```

```
In [18]:
# To drop any duplicates
df['company'].drop_duplicates().sort_values(ascending=False)
```

0	Out[18]:
7225	Micro Scope
4087	15 Films
6793	i an Other
6484	double A Films
3776	Zucker Brothers Productions
3338	Music Services
2213	Zeta Entertainment
2213	Zentropa Entertainments
1138	Zenth Entertainment
5198	Zazen Pictures
3261	Zanuck/Brown Productions
1329	Zacharias-Buhai Productions
6298	Young Sang Productions
5125	Young Hannibal Productions
5499	Yellow Bird
4618	Yassi Film Group
4098	Yari Film Group Releases
5583	Yari Film Group (YFG)
5583	X-Film Creative Pool
6265	X-Wildvision Productions
4999	World of Wonder Productions
4999	World Wrestling Entertainment (WWE)
4272	World Services
450	Working Title Films
3943	Wiseau-Films
2884	WingNut Films
2355	Wildwood Enterprises
6828	Wildwood Enterprises
5276	Wild Bunch
4272	Wiedenmann & Berg Filmproduktion
5559	Why Not Productions
4572	Whitewater Films
1386	WhiteFlame Productions
1315	White Lair
1744	Whitely Films
1384	Weintraub Entertainment Group
5498	Wayfare Entertainment
6672	Wayvly Films
7267	Warner Independent Pictures (WIP)
2284	Warner Bros. Pictures
5198	Warner Bros. Family Entertainment
6198	Warner Bros. Digital Distribution
2341	Warner Bros. Animation
8	Warner Bros. Animation
5484	Warner Animation Group
75	Walt Disney Productions
117	Walt Disney Pictures
5075	Walt Disney Animation Studios
4975	Walden Media
7073	WIP
5272	Voltag Pictures
5489	Vivendi Entertainment
1135	Vista Organization
6633	Visiona Media
3539	Vision PDS
1894	Village Roadshow Pictures
1716	View Askew Productions
4704	Victor & Grais
1532	Victories International
5382	Victrola Entertainment
7018	Versailles
7263	Vendian Entertainment
2584	Vanguard Europe
4287	Vanguard Films and Animation
6288	Visiona Media
7594	Upside Down Films
5035	Universal Pictures International (UPI)
6	Universal Pictures
2943	Universal City Studios
3377	United Artists
227	United Film Distribution Company (UFC)
3817	United Artists Pictures
4686	United Artists Film Corporation
2498	Unison Films
368	Ultra Muchos Productions
4785	UK Film Council
6396	Two Ton Films
4899	Two Prong Lesson
4098	Twisted Pictures
4691	Twentieth Century Fox Animation
281	Twentieth Century Fox
2384	Turner Entertainment Company
6352	Tresor Films
6288	Truth Entertainment (II)
4087	True Entertainment
2369	Trumph Films
3557	Trimark Pictures
2288	Tristar Television
503	Tristar Pictures
6552	Treehouse Pictures
7236	Treehouse Pictures
1290	Trancas International Films
6973	Tony Fight Productions
2881	Touchwood Pacific Partners' LLC
5025	Touchwood Pictures
4691	Total Recall
1845	Tornasoli Films
6473	Torik Productions
4118	Tonic Films
1264	Tonson Films
3356	Tonboy Films
3849	Tokuma Shoten
773	Toku Japan Communications
7432	Toni Pictures
2882	Toni Pictures
1649	Tig Productions
388	Tiberius Film Productions
3828	Three Rivers Production
849	Thorn EMI Screen Entertainment
4312	This Is That Productions
1499	The Zanuck Company
2732	The Weinstein Company
858	The Tyler Perry Company
6398	The Samuel Goldwyn Company
1847	The Safarik Company
1297	The Rank Organisation
10	The Movie Group
5898	The Rank Organisation
2458	The Mark Gordon Company
5855	The Malpaso Company
1339	The Last Picture Company
2524	The Ladd Company
7264	The Lab of Madness
5274	The Kushner-Locke Company
1144	The Imray Company
4233	The Guber-Peters Company
537	The Offbeat Pictures
5824	The American Film Company
5218	The American Film Company
599	Temple Hill Entertainment
4884	Tartan Productions
7459	Taliafilia II Productions
4935	Tag Entertainment
1863	TSD Entertainment
7055	THINKFilm
6285	TACT Entertainment Pictures
2783	Sycamore Pictures
5837	Sweetland Films
4236	Sweet Tea Pictures LLC
2929	Sunbow Productions
2074	Summit Premiere
5388	Studio Canal
7356	Studio 8
5953	Studio 39
4479	Strike Entertainment
4789	Stratus Film Co.
5931	Stone Group Pictures
1883	Steve White Productions
3359	Star Overseas
3954	Stage 6 Films
5837	Stacey Textro International
3483	Spyglass Entertainment
495	Spinal Tap Prod.
7277	SpectreVision
1288	Spectacle Entertainment Group
6841	Sporkia Roll Media
1348	Spacegate Productions
6358	Space Rocket Nation
4123	South Pacific Pictures
4435	Sony Pictures Entertainment (SPE)
1318	Sony Pictures Classics
5114	Sony Pictures Animation
6259	Solana Film
5925	SnooPiercer
4059	Smoot Entertainment
597	Skyscraper Films
7024	Skylark
1978	Siriol Productions
1697	Silver Sphere Corporation
1417	Silver Screen Partners IV
1444	Silver Reel
6498	Silver Pictures
4252	Sho East
3485	Shooting Gallery
5338	Shooting Gallery
7368	ShivHans Productions
3238	Shelberg Productions
6274	Shimada Films
1313	Shapiro-Glickenhau Entertainment
4617	Shangri-La Entertainment
449	Shadowcatcher Entertainment
3742	Seven Arts Pictures
4793	Serendipity Point Films
2757	Seraphim Films
5398	Senator Entertainment Co
4346	SenArt Films
5938	See-Saw Films
5772	Sedic International
374	Sean S. Cunningham Films
3529	Seagull Films
6464	Seagal/Nasso Productions
2946	Screen Gems
7378	Screen Gems
3122	Sciarlo
2583	Savoy Pictures
3768	Saturn Films
2128	Sandollar Productions
276	Sandcastle 5 Productions
3896	Samuelson Productions
1143	Samuel Golden Films
4737	Samuel Golden Films
3282	Samsom Productions Pty. Ltd.
7216	Salon Pictures
7132	Saban Entertainment
3834	SIX Films
7264	SIX Films
6913	SIX Films
6484	SIM Production Group
646	SIS Productions
5985	Sydney Entertainment
2829	Rumblara Films
6625	Ruby Films
5832	Route One Entertainment
6864	Roth Films
4754	Room 9 Entertainment
4522	Rogue Pictures
5511	Rogue
3213	Robert Simonds Productions
6389	Roadside Attractions
2399	River Road Productions
5937	River Road Entertainment
6386	Riviera Films
277	Rifkin-Eberts
3349	Riddick Canada Productions
2388	Richard Light Horse Productions
3349	Richard Williams Productions
3858	Rhombus Media
3785	Revelations Entertainment
1673	Reteitalia
4098	Reprise Pictures
96	Rent Productions LLC
1513	Renn Productions
5167	Renaissance Pictures
6535	Reel FX Creative Studios
3889	Redhawk Films
4891	Red Hour Films
6243	Red Granite Pictures
6236	Red Crown Productions
1733	Recorded Picture Company (RPC)
5767	Realitism Films
129	RafPac-Dune Entertainment
2413	Rastar Productions
15	Rastar Pictures
247	Rastar Pictures
5233	Rankin/Bass Productions
6773	Rainforest Films
2938	Raindog Films
238	Raffaella De Laurentiis Productions
6588	RADIOS-TWC
4186	R.P. Productions
7801	Quinta Communications
3113	Quad Productions
6899	Q Productions
5764	Pure Flix Productions
3924	Punch 21 Productions
3832	P.T. Merantia Films
5656	Profrank Pictures
1288	Protagonist Pictures
861	Producers Sales Organization (PSO)
1497	Producers Representatives Organization
4828	Prime Film Productions LLC
7114	Primate Pictures
2251	Price Entertainment
6485	Pretty Pictures
774	Pressman Productions
5817	Pressman Film
3951	Pressman Pictures
6376	Prana Studios
677	Portman Entertainment Group
3675	Porchlight Pictures
2811	Populit Pictures
2956	Pope Productions
4538	Pop Pop Productions
5452	Portly Pictures
94	Polygram Pictures
2324	Polygram Filmed Entertainment
37	Poly International BV
3388	Polygram Pictures
4384	Polygram Filmed Entertainment
7332	Polar Entertainment
1388	Point Gray Pictures
5171	Plunge Pictures LLC
275	Playtone
6063	Playhouse International
2603	Play It Inc.
6867	Planet Productions
3248	Plan B Entertainment
4349	Pixar Animation Studios
5178	Pioneer LDC
2926	Picturehouse
445	Pictures United
3888	Phoenix Pictures
2438	Phillips Whitehouse Productions
7149	Peters Entertainment
2214	Permut Presentations
2216	Perfect World Pictures
2221	Penta Pictures
5382	Penta Entertainment
7131	Pelican Productions LLC (I)
5685	Pavullu Productions
3517	Pathe Rem Productions
1688	Pathé Pictures International
6386	Pathé Entertainment
2888	Pathé
6651	Party Productions
2378	Parts and Labor
6658	Partisan Productions
2378	Participant
7486	Parkway Productions
5043	Parkes/Lasker productions
2877	Paramount Vantage
6751	Paramount Players
3187	Paramount Pictures
4023	Paramount Pictures
978	Paramount Animation
3187	Paragon Arts International
7286	Papazian-Hirsch Entertainment International
6962	Pantillon Films
2712	Panorama Films
349	Pandora Filmproduktion
5482	Pan European
248	Pan Arts
6659	PalisStar Media
1323	Palla
1328	Palace Pictures
1828	Paisley Park Films
1815	Pacific Western
5198	PWM Entertainment
5597	P2 Productions
5098	Overnight Films
6072	Overbrook Entertainment
1457	Outlaw Productions (I)
4978	Out of the Blue... Entertainment
412	Osternan Weekend Associates
5628	Oscilloscope
6771	Origin Pictures
3353	Oriental Light and Magic (OLM)
5935	Open Road Films (II)
6415	Open Road Films
3752	Open City Films
6813	Onyx Films
159	One Way Productions
5942	Omnilab Media
3944	Omega Project
5795	Omega Entertainment
1835	Olympus Pictures
4193	Odyssey
3021	Odeon Films
5434	October Films
4786	Occupant Entertainment
365	OLC / Rights Entertainment
1763	Number 9 Films
1568	Nova International Films
2134	Northern Lights Films
6339	Northern Lights Entertainment
4998	No Trace Camping
1398	No Frills Film Production
4591	Nine Yards Two Productions
3285	Nimbus Film Productions
5158	Night and Day Pictures
783	Night Life Inc.
1789	Nicolas Entertainment
585	Nibarkil
3318	Next Wave Films
4668	Next Entertainment
526	Newmarket Capital Group
1236	New World Pictures
1346	New World Visions
2737	New Sky Communications Inc.
4828	New Regency Productions
3673	New Oz Productions
2813	New Line Cinema
2414	New Horizons Picture
8164	New Century Entertainment Corporation
3887	New Amsterdam Entertainment
2555	Nest Family Entertainment
485	Nelvana
1885	National Lampoon
6913	Natant
3795	NVSH Productions
3887	NPV Entertainment
2984	NFI Productions
1291	NDF International
3777	NBC Productions
3777	Myung Film Company Ltd.
4943	Mystery Clock Cinema
3993	Myriad Pictures
4127	Muse Productions
2975	Mundy Lane Entertainment
2684	Mulholland Productions
57	Mulberry Square Productions
4886	Movision
6814	Motlys
2977	Motion Picture Corporation of America
5311	Mosaic
4199	Morra, Brezner, Steinberg and Tenebaum Entertainment...
1205	Morgan Creek Productions
7382	Moxy Films
6218	Monsterfoot Productions
7469	Monkeypaw Productions
3481	Monarch Pictures
2432	Mod Productions
681	Miramax
7526	Mirage Enterprises
5829	Miramir Schur Pictures
389	Millennium Media
4778	Milkshake Films
6145	Milk & Media
3928	Milwinter Productions Inc.
3280	Metropolitan Filmexport
3831	Metro-Goldwyn-Mayer (MGM)
4191	Merchant Ivory Productions
6425	Mephisto Films
42	Memento Films Production
5158	Melvin Simon Productions
6553	Medusa Film
5823	Mediaset España
7128	Media Rights Capital (MRC)
4219	Media Asia Films
4267	Media 8 Entertainment
3898	Media8 Entertainment
6984	Mayhem Pictures
6388	Max Films Productions
4225	Matten Productions
6683	Material
6448	Matchbox Pictures
5969	Marvel Studios
6487	Marvel Entertainment
5686	Marv Films
6471	Marked Entertainment
186	Mann/Can Productions
3538	Mandeville Films
5288	Mandate Pictures
5982	Mandalay Vision
4488	Mandalay Pictures
3872	Mandalay Entertainment
114	Maljaco Productions
2238	Major Studio Partners
5574	Majestic Film International
5755	Magnus Pictures Inc.
1292	Magnolia Pictures
6783	Magic Light Pictures
6688	Mace Neufeld Productions
1252	Macari/Edelstein
6783	Mac and Me Joint Venture
669	MTV Films
5993	MPL Communications
6138	MPL Media Group
3731	MK2 Productions
2953	MFF Feature Film Productions
3757	MGM Worldwide
4289	Légende Films
7423	Lydia Films
6578	Lutz-Brown
2599	Lumière Pictures
6248	Lumiere Pictures
368	Lucille Ball Productions
5366	Lucasfilm Animation
7385	Lucasfilm
4884	Lucanar Productions
7151	Lost Soul Productions
286	Los Angeles Media Fund (LAMF)
229	Lorimar Motion Pictures
20	Lorimar Film Entertainment
5958	Longview Entertainment
3981	Livju Productions
2843	Live Entertainment
1982	Little Tokyo Productions
4925	Little Stranger
3348	Lionsgate
1736	Likely Story
5411	Lightyear Entertainment
2699	Lightstone Entertainment
1721	Lightning Pictures
5411	Lightwave Entertainment
3732	Lewitt / Eberts Productions
6359	Levinson Entertainment
3847	Les Films du Losange
1589	Les Films Alain Sande
4175	Les Films 21
226	Les Amateurs
7387	Leisure Investment Company
6487	Legendary Entertainment
6936	LegendEast
7361	LegendSD
652	Legend Production Company
3582	Lee Daniels Entertainment
6855	Lawrence Gordon Productions
1314	Lava Bear Films
1727	Lauren Film
5088	Laurence Mark Productions
2882	Last Holiday Productions
3748	Largo International N.V.
742	Largo Entertainment
1808	Lancaster Gate
5442	Lakeshore Entertainment
371	Ladbrooke
6128	La Petite Reine
6633	Lstan Capital
5531	Leontine Entertainment
7248	LBI Productions
877	L.A. Films
1326	Kuzul Enterprises
1984	Krisjair
2433	Krasnow Productions
6588	Kraner & Sigman Films
2686	Kauf/Riggle Productions
2877	Kopelson Entertainment
3948	Konrad Pictures
4319	Knock Films A.V.V.
4427	Klasky-Csupo
4122	Kingsmeery Properties
929	Kings Road Entertainment
5588	Kin Ki-Duk Film
3934	Killer Films
6887	Kevin Downes Productions
181	Kennedy Miller Productions
2682	Keep Your Head
5387	Kasdan Pictures
4638	Karbo Vantas Entertainment
5104	Kanun-parvareh Fekri
6328	Kang Yu Film Co. Ltd.
3293	K5 International
4838	Kuno Pix
5683	Junegun Movie
1897	Joseph S. Vecchie Entertainment
5281	Jones Entertainment Group
5576	Jon Shestak Productions
6479	John Wells Productions
3573	Jim Henson Pictures
4523	Jet Tone Production
2481	Jersey Films
878	Jerry Bruckheimer Films
3186	Jay Weston Productions
295	Jasmine Productions Inc.
549	James Glickenhaus
1698	Jaffe-Lansing
58	Jackson/McGarry Productions
3488	Jack Rollins & Charles H. Jaffe Productions
3580	Jack Giarraputo Productions
3216	JVC Entertainment Networks
2737	JIT Productions
1465	JF Productions
2737	JOI Productions
5228	J&M Entertainment
1322	Jiclan
5586	It's a Laugh Productions
2082	Island World
917	Island Pictures
1747	Ironbank Films
6883	Iron Horse Entertainment
1769	Intrepid Pictures (II)
528	



2867 Fiction  
5251 Feuldas Entertainment  
7938 Fermin Films  
2867 Fear of God Films  
3256 Fear and Loathing LLC  
2836 Farabi Cinema Foundation  
2578 FallingCloud  
6323 Fairbro House Productions  
6323 Fairview Entertainment  
6905 Fairview Entertainment  
5523 Fade to Black Productions  
5523 Fade To Film  
4161 FX Sound  
2255 FTM Productions  
3288 FTM Productions  
2098 FTM Productions  
1054 FTM Productions  
6111 Exclusive Media Group  
4556 Exclusive Films  
4556 Evolution Entertainment  
4086 Everyman Pictures  
3639 Evenstar Films  
4890 EuropaCorp  
778 Eurasia Investments  
4614 Eulogy Productions LLC  
5618 Euforia Film  
2546 Etalon Film  
4851 Estudios Picasso  
650 Estudios Churubusco Azteca S.A.  
1361 Esperanza Films  
6780 Esparza  
986 Escape Artists  
986 Escalante Productions  
6037 Eros Worldwide  
4203 Epsilon Motion Pictures  
1738 Epic Productions  
112 Eon Productions  
5952 Enticing Entertainment  
7522 Entertainment Studios Motion Pictures  
6764 Entertainment One  
4099 Entertainment Manufacturing Company  
118 Enigma Productions  
453 Englander Productions  
4497 Endgame Entertainment  
2889 Enderby Entertainment  
2089 Encino Man Productions  
1263 Enchanted Pictures  
3432 Enchanter Entertainment  
1108 Enshell Producers  
691 Empire Films  
7084 Empathic Films  
4389 Emperor Multimedia Group (EMG)  
1257 Emmett/Furla/Oasis Films (EFO Films)  
1479 Embassy Pictures  
224 Embassy International Pictures  
3680 Elysian Dreams  
5984 Elio Driver  
2952 Eleventh Street Production  
6656 Element Pictures  
4988 Electric Entertainment  
1452 Electric City Entertainment  
3953 El Deseo  
4825 El Camino Pictures  
4629 Elkon Film  
2519 Egg Pictures  
280 Effer Productions  
4182 Echo Films  
517 Edge City Productions  
604 Edgar J. Scherick Associates  
4216 Eden Media  
1474 Eddie Murphy Productions  
4041 Ecosse Films  
6207 Eclectic Pictures  
5535 Echo Lake Entertainment  
176 Eaves Movie Ranch  
384 Eagle Associates.  
4507 eETIC Films  
195 ERBP  
5296 ERM Films  
6071 EFTI  
5800 Duplass Brothers Productions  
6438 Dune Entertainment  
6438 DreamWorks Pictures  
3278 DreamWorks Animation  
4883 DreamWorks  
2880 Downward  
438 Doug Linan Productions  
7159 Double Nickel Entertainment  
6140 Double Feature Films  
6645 Double Dare You (DDY)  
2128 Dorset Productions  
2869 Donner/Schuler-Donner Productions  
589 Don Simpson/Jerry Bruckheimer Films  
2529 Don Bluth  
3302 Don Roff  
3564 Dogstar Films  
6242 Dog Eat Dog Pictures  
2768 Dog Run Pictures  
7248 Disruption Entertainment  
6636 Disneytoon Studios  
3824 Disney Television Animation  
113 Dino de Laurentiis Communications  
2846 Dimension Films  
2879 Digital Image Associates  
6403 Dickhouse Productions  
7218 Diagonal TV  
3263 DiNovi Pictures  
4765 DI Bonaventura Pictures  
3980 Pharma Productions  
1725 Devoted Productions  
3994 Detour Filmproduction  
2211 Destination Films  
741 Desperate Pictures  
6859 Desert Hearts Productions  
954 Depth of Field  
658 Delphi V Productions  
432 Delphi III Productions  
5234 Delphi Films  
4620 Deep River Productions  
9452 Delphi Films  
168 Delhaven/Shapiro Productions  
5604 De Line Pictures  
410 De Laurentiis Entertainment Group (D&G)  
3151 De Luca-Cappa  
7101 Dayday Films  
3203 Daybreak  
1574 Davros Films  
1899 Davis-Panzer Productions  
2318 Davis-Films  
3199 Davis Entertainment  
779 David Kirschner Productions  
3553 David Foster Productions  
3402 David Brown Productions  
6950 Dark Castle Entertainment  
1472 Danjaq  
4091 Danjaq  
6915 D3 Film  
2311 DIC Entertainment  
6019 DENTSU Music And Entertainment  
3149 DEI Productions  
444 D C Entertainment  
5551 D & P Productions  
782 DCP Productions  
3559 Crystalite Worldwde  
3559 Crystal Sky Worldwide  
3389 Crunk Pictures  
3722 Cruise/Wagner Productions  
2456 Cruella Productions  
3695 Crowission Inc.  
5907 Crossroads Films  
5969 Cross Creek Pictures  
1257 Cristallidfilm  
4822 Crescent Moon  
7098 Crescent Drive Pictures  
3609 Cre Film  
3695 Craven-Maddalena Films  
6807 Cradle Productions Inc.  
5907 Covert Media  
3083 Cottonwood Pictures  
5969 Cort/Madden Pictures  
1324 Corsan  
3681 Cornelius Productions  
4357 Copperheart Entertainment  
2702 Contentfile  
3279 Constellation Entertainment  
487 Constantin Film  
2151 Consejo Nacional para la Cultura y las Artes (...  
1194 Conserving Unicorn  
1717 Concorde Pictures  
3678 Compulsion Inc.  
5827 Company Films  
1285 Comies From Mars Corporation  
1180 Columbia Pictures Industries  
4516 Columbia Pictures Film Production Asia  
5702 Color Force  
536 Coligems Productions Ltd.  
5437 Cohen Media Group  
6575 Codeblack Entertainment  
3736 Code Entertainment  
6238 Codetail Film  
385 Cobalt Media Group  
4697 Coach Carter  
7645 Clubhouse Pictures (II)  
3771 Cloud Ten Pictures  
3745 Cloud Eight Films  
6688 Cloud Atlas Productions  
5689 Closest to the Hole Productions  
3655 Clispal Films  
3197 Clinica Estetico  
5945 Cliffjack Motion Pictures  
282 Clavius Base  
3806 Claude Ossard Productions  
1062 City Films  
3395 Cineplex Film Properties (CFP)  
2614 Cineville  
1619 Cinetel Films  
538 Cinesesthesia Productions  
2493 Cinerigi Pictures  
3789 Cinerenta Medienbeteiligungs KG  
6187 Cinerach  
718 Cinepro  
1212 Cineplex Odeon Films  
231 Cinematograph AB  
1800 Cinenarock Entertainment  
3628 Cinenarock  
3084 Cinema Line Film Corporation  
180 Cinema Group Ventures  
1317 Cinema City Films  
187 Cinema "84  
6587 Cinelou Films  
1712 Cineplex  
1486 Cinecorp SMC  
1146 Cinecom Entertainment Group  
4231 CineWorld  
5616 CineSon Entertainment  
1604 Cine Location Services  
2269 CIBY 2000  
2195 Christopher Columbus Productions  
7455 Chris Morgan Productions  
3965 Chris Lee Productions  
6337 Clockstone Pictures  
2342 Clodo Brothers Productions  
3426 Chinese Bookie Pictures  
6944 China Film Group Corporation (CFG)  
3189 China Film Co-Production Corporation  
4148 Chickie the Cop  
4148 Cheyenne Enterprises  
7229 Chestnut Ridge Productions  
3745 Chester Films Inc.  
4134 Cherry Sky Films  
3633 Cherry Alley Productions  
6307 Chemin Entertainment  
3403 Cheerleader LLC  
6212 Chase Films  
7 Chertoff-Winkler Productions  
99 Charles H. Schneer Productions  
467 Charles Burnett Productions  
2173 Channel Four Films  
2576 Channel Films  
3686 Challenge Film Corporation  
2673 Chal Productions  
4771 Centurion  
3233 Centropolis Film Productions  
4771 Cent Productions Inc.  
5077 Celluloid Dreams  
5877 Celandine Films  
5266 Cellador Films  
2645 Cecchi Gori Pictures  
1401 Cecchi Gori Group  
2400 Cecchi Gori Europa N.V.  
6759 Caviar  
33 Caw 76  
4299 Cattle Annie Productions  
1458 Casual Friday Productions  
4213 Castle Rock Entertainment  
4213 Castello Producciones  
5956 Casey Silver Productions  
5138 Carver Films  
9326 Carthago Films S.a.r.l.  
5138 Carsey-Werner Company  
5858 Carousel Productions (II)  
1145 Carolco Pictures  
1805 Carolco International N.V.  
1504 Carolco Entertainment  
1976 Carnival Film & Television  
2477 Caravan Pictures  
5208 Captivity Productions  
569 Capricorn  
3840 Cappa Production  
5116 Capital Films  
932 Capital Equipment Leasing  
1936 Capella International  
2373 Capella Films  
1130 Capcon Entertainment  
406 Canon Italia S.r.l.  
393 Canon Pictures  
1305 Canon International  
2045 Candyman Films  
7557 Canana Films  
4578 Canal+ España  
3546 Canal+ Droits Audiovisuels  
2094 Canal+  
1566 Canadian International Studios VIII  
37 Camp Hill  
4518 Camelot Pictures  
228 Cambridge Productions  
6375 Callahan Filmworks  
3298 Calinari Productions  
7457 CalMaple  
312 Cable and Wireless Finance  
398 Cabal Film  
5215 CTB Film Company  
2319 CNCMC  
4128 C3 Entertainment  
134 CIP Filmproduction GmbH  
7015 CO Cinema  
611 CBS Theatrical Films  
5722 CBS Films  
512 C.O.R.E. Feature Animation  
11 C.M.U.D. Productions  
4265 C-2 Films  
305 C & C Brown Production  
5154 Córd Scannán na hÍreann  
3192 Butcher's Run Films  
6923 Burk A Project  
3017 Bungalow Productions  
4603 Bullwinkle Studios  
6980 Bull's Eye Entertainment  
4836 Bulbul Films  
7477 Buena Vista International  
4462 Buena Vista Home Entertainment  
795 Bud Yorkin Productions  
2550 Buckeye Films  
22 Bryna Productions  
2962 Bruin Grip Services  
6648 Brownstone Productions (II)  
7095 Brownstone Entertainment (II)  
2794 Brothers McMullen Productions  
1568 Brooksfilms  
5638 Bronze Eye Productions  
6529 Broken Road Productions  
4576 Broken Lizard Industries  
5369 Broadway Video  
3327 Broadway Pictures  
6638 Broad Green Pictures  
1215 British Screen Productions  
3362 British Broadcasting Corporation (BBC) Television  
2149 British Broadcasting Corporation (BBC)  
3025 Brillstein-Grey Entertainment  
1785 Briland Films  
5415 Bridgit Folman Film Gang  
363 Breathless Associates  
7070 Branti Film Productions  
3758 Brant-Allen  
3072 Brandywine Productions  
6983 Bow and Arrow Entertainment  
88 Borough Park Productions  
7014 Bona Film Group  
3473 Bona Fide Productions  
4342 Boll Kino Beteiligungs GmbH & Co. KG  
5028 Bold Film  
7262 Boies / Schiller Film Group  
4477 Bob Yari Productions  
6501 Bob Industries  
6186 Blumhouse Productions  
5934 Blueprint Pictures  
3682 Bluegrass Films  
6230 Blue Vonder Films  
3816 Blue Tulip Productions  
4157 Blue Train Productions  
3421 Blue Scream Films  
5339 Blue Sky Studios  
6943 Blue Sky Films  
1318 Blue Rider Pictures  
7313 Blue Budgie Films Limited  
5259 Blue Askey  
4033 Blow Up Pictures  
6888 Blown  
4187 Blockbuster Films  
3114 Block 2 Pictures  
6158 Block / Hanson  
6279 Blinding Edge Pictures  
6119 Bleiberg Entertainment  
6733 Bleecker Street Films  
5598 Blank of the Dead Productions  
6639 BlackWhite  
6765 Black Label Media  
5523 Black Entertainment Television (BET)  
7054 Bicycle Entertainment  
6461 Black Bear Pictures  
5639 Bill Kenwright Films  
1854 Bill Graham Films  
3613 Bigel / Mailer Films  
7566 Big Talk Productions  
6493 Big Screen Productions  
4640 Big Red Dog Productions  
6834 Big Indie Pictures  
3354 Big Dog Films  
5418 Big City Pictures  
6083 Big Beach Films  
4078 Beverly Detroit  
4744 Berk/Lane Entertainment  
4468 Bergman Lustig Productions  
2589 BenderSpink  
6209 Ben-Ami/Leeds Productions  
4821 Belladonna Productions  
3703 Bel Air Entertainment  
6249 Beijing Skywheel Entertainment Co.  
6923 Beijing New Picture Film  
7652 Beijing Digi Vinxlang Entertainment  
7565 Beijing Dengfeng International Culture Commu...  
5915 Before The Door Pictures  
1586 Beco Films  
2700 Beckner/Gorman Productions  
1804 Beacon Communications  
2845 Bazmark Films  
4605 Bazelevs Productions  
5668 Bavaria Film  
3578 Baumgarten-Prophet Entertainment  
3218 Bates Entertainment  
5398 Barunson E&A  
133 Barry & Enright Productions  
115 Barclays Mercantile Industrial Finance  
2624 Barcelona Films  
3791 Banner Entertainment  
3210 Bandeira Entertainment  
2311 Bandai Visual Company  
1968 Baltimore Pictures  
3558 Baldwin/Cohen Productions  
740 Balcon Film Investors  
2866 Bakshi Animation  
5749 Bad Robot  
2096 Bad Lt. Productions  
6132 Bad Cop Bad Cop Film Productions  
2951 Bad Bird Productions  
504 Bachelor Party Productions  
5387 Babylon  
6585 Baby Face Productions  
6895 B2 Entertainment  
6687 BRON Studios  
6826 BN Films  
4748 BET Pictures  
5139 BET Films  
6201 BCDF Pictures  
6480 BBL Motion Picture Studios  
594 BBG  
2825 BBC Films  
6665 B Story  
7123 B Story  
3530 B Story  
7143 B Story  
3359 Award Entertainment  
3274 Avrono Media  
2855 Avnet/Kerner Productions  
7333 Aviron Pictures  
7424 Avi Arad Productions  
5168 Avi Arad & Associates  
3786 Avery Pix  
1508 Avenue Pictures  
5625 Autonomous Films  
6563 Automatix Entertainment  
2932 Australian Film Production Corporation (APFC)  
7564 Aurora Alliance Films  
223 Aurora  
2371 August Entertainment  
3387 Audiovisual Development Bureau, Ministerio da ...  
5705 Audiovisual Aval SGR  
6871 Audifilm  
2822 Audax Films  
7285 Atomic Monster  
6027 Atlas Productions  
3311 Atlas Entertainment  
7558 Atlantic Records  
6157 Atlantic Pictures (II)  
2189 Atchafalaya  
3510 Asymmetrical Productions  
6981 Astute Films  
3474 Astralwerks  
975 Astral Believe Films  
6935 Astrakhan Film  
2847 Astoria Films  
258 Aspen Film Society  
3675 Asia Union Film & Entertainment Ltd.  
5103 Asghar Farhadi Productions  
2943 Ascot Film  
5390 Ascendant Pictures  
3596 Arts Council of England  
5970 Artists Public Domain  
1481 Artists First  
509 Artistry Limited  
6141 Artista Management  
3423 Artisan Entertainment  
2364 Artisan Entertainment  
3755 Artistic Productions LLC  
5213 Artfire Films  
7154 Artbees  
1479 Art Linson Productions  
1914 Arnold Kopelson Productions  
3760 Armony Films  
4914 Arma MediaWorks  
3790 Aries Cinematográfica Argentina  
4798 AriesScope Pictures  
1935 Argos Film  
1961 Arena  
4386 Archer Street Productions  
7377 Archer Gray  
529 Arcatin B.V.  
6171 Arcade Pictures  
5031 Arc Productions  
5838 Aramid Entertainment Fund  
6298 Applied Action  
3886 Applan Way  
4212 ApollMedia Distribution  
5128 Apatow Productions  
7348 Apartment Story  
5225 Anonymous Content  
2635 Annhall  
5840 Annapurna Productions  
7139 Annapurna Pictures  
3804 Animal Productions LLC  
6219 Animal Logic  
7139 Animal Kingdom  
1852 Anheilo Producciones  
4298 Angry Films  
2405 Ang Lee Productions  
6355 Andrew Lauren Productions  
4701 Andrea Sperling Productions  
5911 Anchor Bay Films  
5900 Anarchist's Convention Films  
218 Anshah N.V.  
2048 American Zoetrope  
1627 American Playhouse Theatrical Films  
714 American Playhouse  
977 American Filmworks  
1246 American Entertainment Partners II L.P.  
3268 American Empirical Pictures  
1804 American Cinema Productions  
1084 Amercent Films  
3252 Anen Ra Films  
3575 Ambridge Film Partnership  
6893 Amblin Entertainment  
6738 Amazon Studios  
3644 Am Psycho Productions  
4610 Alter Ego Entertainment  
3650 Altavista Films  
6180 Also Known As Pictures  
4186 Alphaville Films  
639 Almost You  
6824 Alton Films  
7131 Allspark Pictures  
6602 Alloy Entertainment  
6602 Allison Shearman Productions  
1525 Allied Vision  
2758 Allied Stars Ltd.  
2876 Allied FilmMakers  
6186 Alliance Communications  
3233 Alliance Atlantis Communications  
1529 Alleged Productions  
3424 Allan Zeman Productions  
3203 All Net Productions  
1204 All Girl Productions  
1889 Alive Films  
7385 Alibaba Pictures  
6459 Afrana Films  
2172 Aldamisa Entertainment  
6255 Alcor Films  
2625 Alcon Entertainment  
6271 Aim  
3212 Aggregate Films  
6482 Agenda  
7537 After-Dark Films  
1914 Affirm Films  
2922 Adventure Pictures  
5911 Addis Meschler Pictures  
392 Adams Apple Film Company  
6234 Ada Films  
1643 AdScott Pictures  
7099 Act III Communications  
7079 Access Entertainment  
1384 Acacia Filmed Entertainment  
6091 Abramoff Production  
6370 Abraham Productions  
5324 Abbotella Productions  
1129 Abandon Entertainment  
3667 Aaron Spelling Productions  
6218 Aardman Animations  
18 Aamir Khan Productions  
6584 AXVO Embassy Pictures  
484 ASIG Productions  
1805 AR Films  
4802 AM/PM Entertainment  
1781 AITD Productions  
367 ABC Motion Pictures  
6452 A24  
544 A&M Films  
2819 A Band Apart  
1456 98 MPH Productions  
4590 900 Films  
1655 888 Productions  
1690 88 Productions  
3829 7 Films Cinema  
3083 7 Arts International  
5161 4Kids Entertainment  
913 40 Acres & A Mule Filmworks  
3592 4 Kids Entertainment  
4038 3Mark Entertainment  
5199 360 Pictures  
7213 3111 Productions  
3279 3 Miles Apart Productions Ltd.  
7005 3 Arts Entertainment  
5218 2DUX  
5597 2029 Productions  
4118 26 Films  
2522 25th Hour Productions  
6528 21st Century Film Corporation  
7651 21 Laps Entertainment  
4559 20th Century Studios  
5195 2003 Productions  
7189 2.4.7. Films  
4412 19 Entertainment  
385 1818 Productions  
2929 1492 Pictures  
3024 486 Production  
7525 "Weathering With You" Film Partners  
4345 "DIA" Productions GmbH & Co. KG  
633 NaN  
Name: company, dtype: object





Correlation Matric for Numeric Features



```
In [37]:
df.head()

Out[37]:
```

	name	rating	genre	year	released	score	votes	director	writer	star	country	budget
5445	Avatar	PG-13	Action	2009	December 18, 2009 (United States)	7.8	1100000.0	James Cameron	James Cameron	Sam Worthington	United States	237000000 2847.
7445	Avengers: Endgame	PG-13	Action	2019	April 26, 2019 (United States)	8.4	903000.0	Anthony Russo	Christopher Markus	Robert Downey Jr.	United States	356000000 2797.
3045	Titanic	PG-13	Drama	1997	December 19, 1997 (United States)	7.8	1100000.0	James Cameron	James Cameron	Leonardo DiCaprio	United States	200000000 2201.
6663	Star Wars: Episode VII - The Force Awakens	PG-13	Action	2015	December 18, 2015 (United States)	7.8	876000.0	J.J. Abrams	Lawrence Kasdan	Daisy Ridley	United States	245000000 2069.
7244	Avengers: Infinity War	PG-13	Action	2018	April 27, 2018 (United States)	8.4	897000.0	Anthony Russo	Christopher Markus	Robert Downey Jr.	United States	321000000 2048.

```
In [15]:
df_numerized = df
```

```
In [31]:
for col_name in df.numerized.columns:
    if(df.numerized[col_name].dtype == 'object'):
        df_numerized[col_name] = df_numerized[col_name].astype('category')
        df_numerized[col_name] = df_numerized[col_name].cat.codes

df_numerized.head()

Out[31]:
```

	name	rating	genre	year	released	score	votes	director	writer	star	country	budget	gross	company	n
0	6587	6	6	1980	1705	8.4	927000.0	2589	4014	1047	54	19000000.0	46998772.0		2319
1	5573	6	1	1980	1492	5.8	650000.0	2269	1632	327	55	45000000.0	58853106.0		731
2	5142	4	0	1980	1771	8.7	1200000.0	1111	2567	1745	55	18000000.0	538375067.0		1540
3	286	4	4	1980	1492	7.7	221000.0	1301	2000	2246	55	3500000.0	83453539.0		1812
4	1027	6	4	1980	1543	7.3	1080000.0	1054	521	410	55	60000000.0	39846344.0		1777

```
In [19]:
correlation_matrix = df_numerized.corr()

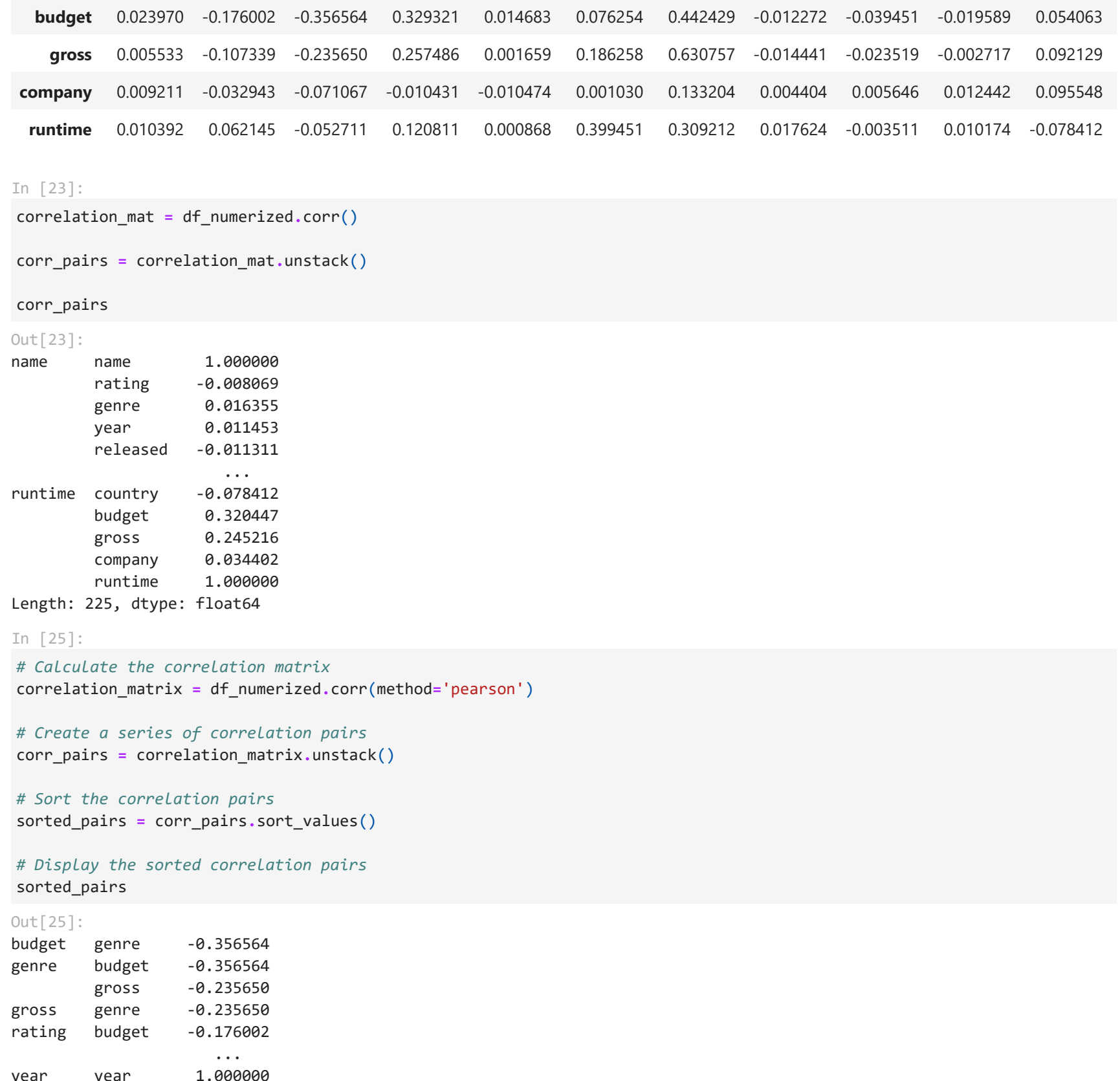
sns.heatmap(correlation_matrix, annot= True)

plt.title('Correlation Matric for Numeric Features')

plt.xlabel('Movie Features')

plt.ylabel('Movie Features')

plt.show()
```



```
In [28]:
df_numerized.corr()

Out[28]:
```

	name	rating	genre	year	released	score	votes	director	writer	star	country
name	1.000000	-0.008069	0.016355	0.011453	-0.011311	0.017097	0.013088	0.009079	0.009081	0.006472	-0.010737
rating	-0.008069	1.000000	0.072423	0.008779	0.016613	-0.001314	0.033225	0.019483	-0.005921	0.013405	0.081244
genre	0.016355	0.072423	1.000000	-0.081261	0.029822	0.027965	-0.145307	0.015258	0.006567	-0.005477	-0.037615
year	0.011453	0.008779	-0.081261	1.000000	-0.000695	0.097995	0.222945	-0.020795	-0.008656	-0.027242	-0.070938
released	-0.011311	0.016613	0.029822	-0.000695	1.000000	0.042788	0.016097	-0.001478	-0.002404	0.015777	-0.020427
score	0.017097	-0.001314	0.027965	0.097995	0.042788	1.000000	0.409182	0.009559	0.019416	-0.001609	-0.133348
votes	0.013088	0.033225	-0.145307	0.222945	0.016097	0.409182	1.000000	0.000260	0.000892	-0.019282	0.073625
director	0.009079	0.015258	-0.005477	-0.020795	-0.001478	0.009559	0.000260	1.000000	0.299067	0.039234	0.017490
writer	0.009081	-0.005921	0.006567	-0.008656	-0.002404	0.019416	0.000892	0.299067	1.000000	0.027245	0.015343
star	0.006472	0.013405	-0.005477	-0.027242	0.015777	-0.001609	-0.019282	0.039234	0.027245	1.000000	-0.012998
country	-0.010737	0.081244	-0.037615	-0.070938	-0.020427	-0.133348	0.073625	0.017490	0.015343	-0.012998	1.000000
budget	0.023970	-0.176002	-0.356564	0.329321	0.014683	0.076254	0.442429	-0.012272	-0.039451	-0.019589	0.054063
gross	0.005533	-0.110739	-0.235650	0.257486	0.001659	0.186258	0.630757	-0.014441	-0.023519	-0.022717	0.092129
company	0.009211	-0.032943	-0.071067	-0.010431	-0.010474	0.001030	0.133204	0.004404	0.005646	0.012442	0.095548
runtime	0.010392	0.062145	-0.052711	0.120811	0.000868	0.399451	0.309212	0.017624	-0.003511	0.010174	-0.078412

```
In [23]:
correlation_mat = df_numerized.corr()

corr_pairs = correlation_mat.unstack()

Out[23]:
```

name	name	1.000000
rating	rating	-0.008069
genre	genre	0.016355
year	year	0.011453
released	released	-0.011311
score	score	0.017097
votes	votes	0.013088
director	director	0.009079
writer	writer	0.009081
star	star	0.006472
country	country	-0.010737
budget	budget	0.023970
gross	gross	0.005533
company	company	0.009211
runtime	runtime	0.010392

```
Length: 225, dtype: float64

In [25]:
# Calculate the correlation matrix
correlation_matrix = df_numerized.corr(method='pearson')

# Create a series of correlation pairs
corr_pairs = correlation_matrix.unstack()

# Sort the correlation pairs
sorted_pairs = corr_pairs.sort_values()

# Display the sorted correlation pairs
sorted_pairs
```

budget	genre	-0.356564
budget	rating	-0.356564
gross	genre	-0.235650
gross	rating	-0.235650
rating	budget	-0.176002
year	year	1.000000
genre	genre	1.000000
rating	rating	1.000000
company	company	1.000000
runtime	runtime	1.000000

```
Length: 225, dtype: float64

In [28]:
# Calculate the correlation matrix
correlation_matrix = df_numerized.corr(method='pearson')

# Create a series of correlation pairs
corr_pairs = correlation_matrix.unstack()

# Sort the correlation pairs
sorted_pairs = corr_pairs.sort_values()

# Filter for high correlations
high_corr = sorted_pairs(sorted_pairs) > 0.5

# Display the high correlations
print(high_corr)
```

gross	votes	0.638757
votes	gross	0.638757
budget	gross	0.746395
gross	budget	0.746395
name	name	1.000000
director	director	1.000000
gross	gross	1.000000
budget	budget	1.000000
country	country	1.000000
star	star	1.000000
writer	writer	1.000000
writer	writer	1.000000
votes	votes	1.000000
score	score	1.000000
released	released	1.000000
year	year	1.000000
genre	genre	1.000000
rating	rating	1.000000
company	company	1.000000
runtime	runtime	1.000000

```
dtype: float64

In [ ]:
# Votes and budget have the highest correlation to gross earnings
#company has low correlation
```