

Визуализация внешних данных

Письменный Алексей

30 сентября 2019 г.

Данные

- Взяты с <https://www.kaggle.com/danielgrijalvas/movies>
- Данные по самым популярным фильмам с 1986 по 2016 год (по 220 фильмов в год), с IMDB

	budget	company	country	director	genre	gross	name	rating	released	runtime	score	star	votes	writer	year
0	8000000.0	Columbia Pictures Corporation	USA	Rob Reiner	Adventure	52287414.0	Stand by Me	R	1986-08-22	89	8.1	Wil Wheaton	299174	Stephen King	1986
1	6000000.0	Paramount Pictures	USA	John Hughes	Comedy	70136369.0	Ferris Bueller's Day Off	PG-13	1986-06-11	103	7.8	Matthew Broderick	264740	John Hughes	1986
2	15000000.0	Paramount Pictures	USA	Tony Scott	Action	179800601.0	Top Gun	PG	1986-05-16	110	6.9	Tom Cruise	236909	Jim Cash	1986
3	18500000.0	Twentieth Century Fox Film Corporation	USA	James Cameron	Action	85160248.0	Aliens	R	1986-07-18	137	8.4	Sigourney Weaver	540152	James Cameron	1986
4	9000000.0	Walt Disney Pictures	USA	Randal Kleiser	Adventure	18564613.0	Flight of the Navigator	PG	1986-08-01	90	6.9	Joey Cramer	36636	Mark H. Baker	1986
5	6000000.0	Hemdale	UK	Oliver Stone	Drama	138530565.0	Platoon	R	1987-02-06	120	8.1	Charlie Sheen	317585	Oliver Stone	1986
6	25000000.0	Henson Associates (HA)	UK	Jim Henson	Adventure	12729917.0	Labyrinth	PG	1986-06-27	101	7.4	David Bowie	102879	Dennis Lee	1986

Данные

- 6820 фильма
- 57 стран
- 2759 режиссера
- 4199 сценаристов
- 17 жанров
- 2179 компании
- 2504 главных актеров/актрис

Пропуски / ошибки

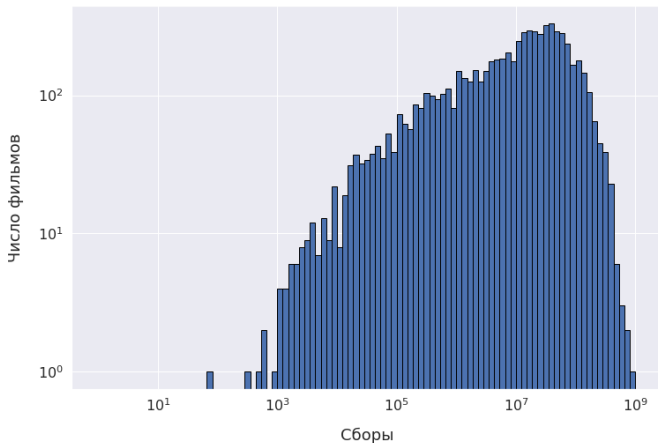
- 2182 фильма с «нулевым» бюджетом (32%)
- В описании датасета указано, что это пропуски

	budget	company	country	director	genre	gross	name	rating
25	0.0	TriStar Pictures	USA	John Badham	Comedy	40697761.0	Short Circuit	PG
26	0.0	Neue Constantin Film	Italy	Jean-Jacques Annaud	Crime	7153487.0	The Name of the Rose	R
27	0.0	TriStar Pictures	USA	Sidney J. Furie	Action	24159872.0	Iron Eagle	PG-13
32	0.0	Gaumont	France	Jean-Jacques Beineix	Drama	2003822.0	Betty Blue	Not specified
35	0.0	Columbia Pictures Corporation	USA	John G. Avildsen	Action	115103979.0	The Karate Kid Part II	PG
36	0.0	Lorimar Productions	USA	Nick Castle	Drama	7177431.0	The Boy Who Could Fly	PG
41	0.0	Universal Pictures	USA	Don Bluth	Animation	47483002.0	An American Tail	G
48	0.0	Metro-Goldwyn-Mayer (MGM)	USA	Peter Hyams	Action	38500726.0	Running Scared	R
49	0.0	Hawn / Sylbert Movie Company	USA	Michael Ritchie	Comedy	26285544.0	Wildcats	R

Пропуски / ошибки

- Минимальные сборы фильма (по данным) - 70\$ (при «нулевом») бюджете
- Его сборы по данным Википедии - \$10.9K
- Кажется, дело не ограничивается нулями, смотрим дальше

Статистика сборов



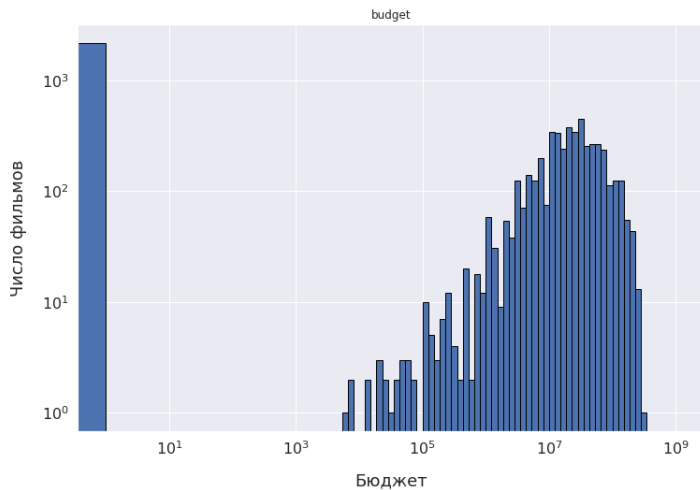
- Буквально несколько фильмов собрали меньше 1000\$

Статистика сборов

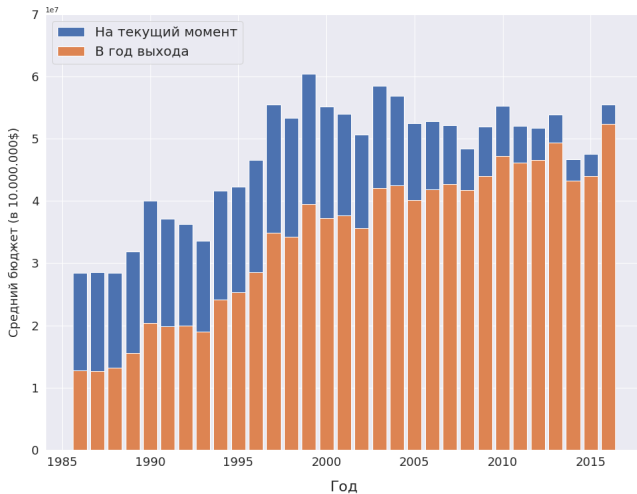
	budget	name	director	gross	country
6784	0.0	Pet	Carles Torrens	70.0	Spain
2594	15000000.0	Trojan War	George Huang	309.0	USA
6066	0.0	I Spit on Your Grave 2	Steven R. Monroe	441.0	USA
5699	0.0	The Decoy Bride	Sheree Folkson	542.0	UK
5917	0.0	Into the White	Petter Næss	547.0	Norway
5242	0.0	Perrier's Bounty	Ian Fitzgibbon	828.0	Ireland
1319	0.0	Samantha	Stephen La Rocque	1076.0	USA
5711	5000000.0	Flypaper	Rob Minkoff	1098.0	Germany
3255	0.0	Love, Honor and Obey	Dominic Anciano	1143.0	UK
5671	0.0	Chalet Girl	Phil Traill	1201.0	UK

- С первым фильмом скорее всего вышла ошибка
- В других случаях указаны (правильно) сборы в США

Статистика по бюджету



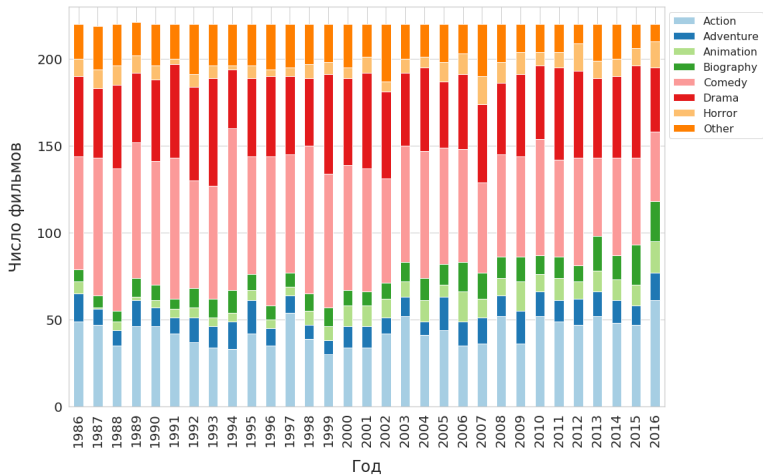
Статистика по бюджету



Жанры

genre	Action	Adventure	Animation	Biography	Comedy	Crime	Drama	Family	Fantasy	Horror	Musical	Mystery	Romance	Sci-Fi	Thriller	War
year																
1986	49	16	7	7	65	12	46	2	0	10	1	0	0	1	4	0
1987	47	9	1	7	79	17	40	0	4	11	0	1	0	1	2	0
1988	35	9	5	6	82	14	48	1	3	11	0	1	2	0	1	2
1989	46	15	2	11	78	13	40	1	3	10	0	0	0	1	1	0
1990	46	11	4	9	71	17	47	2	2	8	0	1	0	1	1	0
1991	42	9	5	6	81	15	54	0	0	3	0	3	1	0	1	0
1992	37	14	6	11	62	25	54	0	1	7	0	1	1	0	1	0
1993	34	12	5	11	65	19	62	1	2	7	0	0	1	0	0	0
1994	33	16	5	13	93	17	34	3	2	2	0	1	1	0	0	0
1995	42	19	6	9	68	18	45	1	2	7	1	1	0	1	0	0
1996	35	10	5	8	86	19	46	2	1	4	0	1	1	1	1	0
1997	54	10	5	8	68	19	45	0	3	5	0	1	2	0	0	0
1998	39	8	8	10	85	20	39	0	0	8	0	1	0	0	2	0
1999	30	8	8	11	77	15	57	0	3	7	1	3	0	0	0	0
2000	34	12	12	9	72	22	50	0	0	6	0	1	2	0	0	0
2001	34	12	12	8	71	17	55	0	1	9	0	1	0	0	0	0
2002	42	9	11	9	60	29	50	0	0	6	1	1	2	0	0	0
2003	52	11	9	11	67	17	42	0	0	8	0	3	0	0	0	0
2004	41	8	12	13	73	16	48	1	0	6	0	1	0	1	0	0
2005	44	19	7	12	67	21	38	0	0	11	0	1	0	0	0	0
2006	35	14	17	17	65	16	43	0	0	12	0	1	0	0	0	0
2007	36	15	11	15	52	27	45	0	0	16	0	2	0	1	0	0
2008	52	12	10	12	59	21	41	0	1	12	0	0	0	0	0	0
2009	36	19	17	14	58	12	47	0	0	13	0	2	0	1	1	0
2010	52	14	10	11	67	11	42	0	1	8	0	1	0	1	2	0
2011	49	12	13	12	56	10	53	0	2	9	0	3	1	0	0	0
2012	47	15	10	9	62	9	50	0	0	16	0	1	0	1	0	0
2013	52	14	12	20	45	18	46	0	1	10	0	2	0	0	0	0
2014	48	13	12	14	56	16	47	0	0	10	0	2	1	1	0	0
2015	47	11	12	23	50	10	53	0	0	10	0	1	0	1	1	0
2016	61	16	18	23	40	10	37	0	0	15	0	0	0	0	0	0

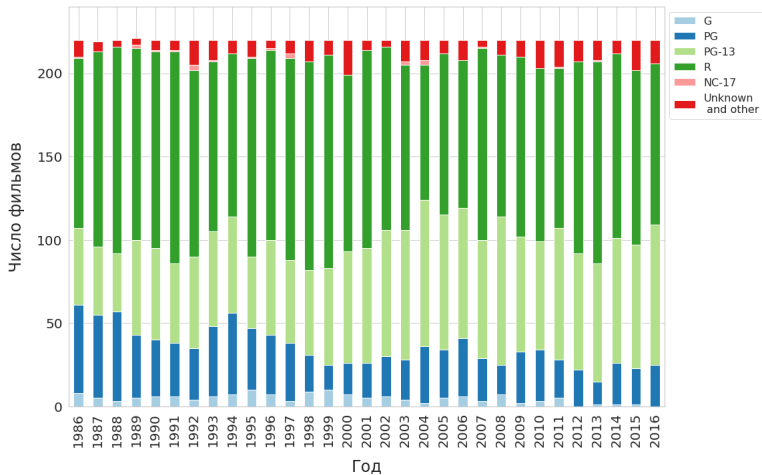
Жанры



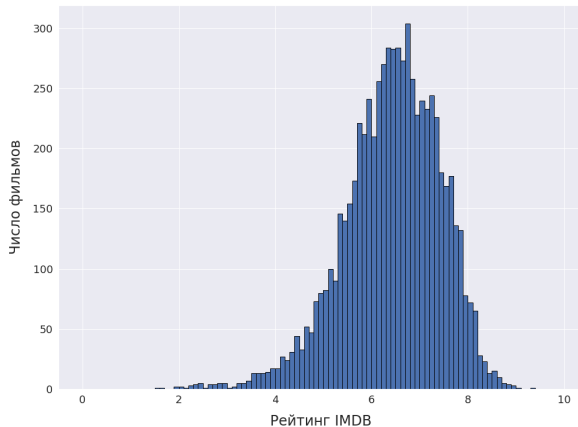
Возрастные рейтинги

rating	G	PG	PG-13	R	NC-17	NOT RATED	UNRATED	Not specified	TV-PG	TV-MA	B15	TV-14	B
year													
1986	8	53	46	102	1	0	4	6	0	0	0	0	0
1987	5	50	41	117	0	2	0	4	0	0	0	0	0
1988	3	54	35	124	0	1	0	3	0	0	0	0	0
1989	5	38	57	115	2	2	0	2	0	0	0	0	0
1990	6	34	55	118	1	2	1	3	0	0	0	0	0
1991	6	32	48	127	1	2	2	1	1	0	0	0	0
1992	4	31	55	112	3	3	6	6	0	0	0	0	0
1993	6	42	57	102	1	5	3	4	0	0	0	0	0
1994	7	49	58	98	0	5	3	0	0	0	0	0	0
1995	10	37	43	119	1	5	4	0	0	1	0	0	0
1996	7	36	57	114	1	4	0	1	0	0	0	0	0
1997	3	35	50	121	3	3	2	3	0	0	0	0	0
1998	9	22	51	125	0	4	7	2	0	0	0	0	0
1999	10	15	58	128	0	4	1	4	0	0	0	0	0
2000	7	19	67	106	0	7	7	6	0	0	0	0	1
2001	5	21	69	119	0	4	1	1	0	0	0	0	0
2002	6	24	76	110	0	2	1	1	0	0	0	0	0
2003	4	24	78	99	2	5	4	3	0	0	1	0	0
2004	2	34	88	81	3	8	2	2	0	0	0	0	0
2005	5	29	81	97	0	4	4	0	0	0	0	0	0
2006	6	35	78	89	0	9	3	0	0	0	0	0	0
2007	3	26	71	115	1	2	2	0	0	0	0	0	0
2008	7	18	89	97	0	8	0	1	0	0	0	0	0
2009	2	31	69	108	0	7	1	2	0	0	0	0	0
2010	3	31	65	104	0	13	3	1	0	0	0	0	0
2011	5	23	79	96	1	14	1	1	0	0	0	0	0
2012	0	22	70	115	0	12	0	1	0	0	0	0	0
2013	1	14	71	121	1	11	1	0	0	0	0	0	0
2014	1	25	75	111	0	5	2	1	0	0	0	0	0
2015	1	22	74	105	0	13	1	3	0	0	0	1	0
2016	0	25	84	97	0	8	5	1	0	0	0	0	0

Возрастные рейтинги

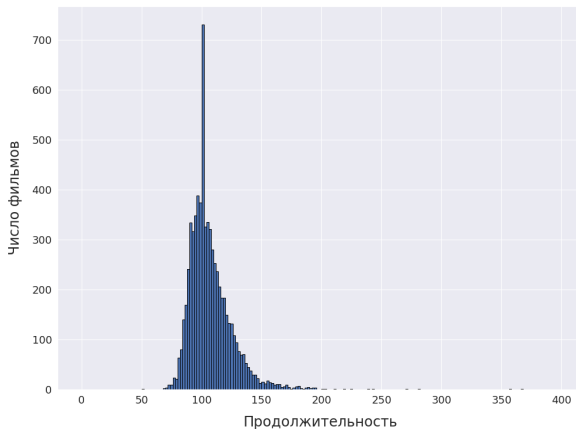


Рейтинг пользователей



- Рейтинг определяется с точностью до 0.1

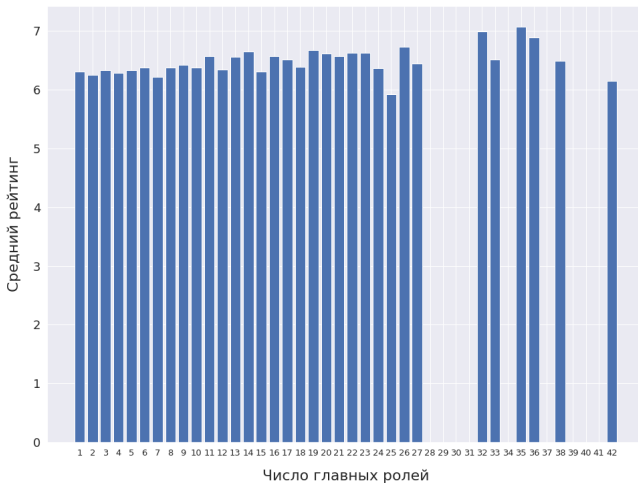
Продолжительность фильма



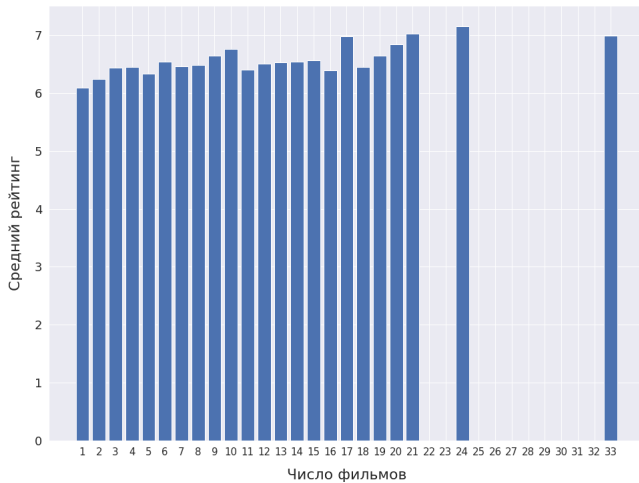
- 546 фильмов длиной 100 минут и 185 длиной в 101 минуту
- Со 100 минутами вышла ошибка

Насколько коррелируют рейтинг и популярность главного актера?

- Для каждого главного актера смотрим на число главных ролей



Или режиссера?



Топ-25 режиссеров

	Максимальный рейтинг	Средний рейтинг
0	Frank Darabont 9.3	Nitesh Tiwari 8.7
1	Christopher Nolan 9.0	Anurag Kashyap 8.6
2	Steven Spielberg 8.9	Olivier Nakache 8.6
3	Quentin Tarantino 8.9	Marco Tullio Giordana 8.5
4	Peter Jackson 8.9	Aamir Khan 8.5
5	Robert Zemeckis 8.8	Makoto Shinkai 8.5
6	David Fincher 8.8	Nishikant Kamat 8.4
7	Martin Scorsese 8.7	Damien Chazelle 8.35
8	Lana Wachowski 8.7	Christopher Nolan 8.322
9	Fernando Meirelles 8.7	Majid Majidi 8.3
10	Nitesh Tiwari 8.7	Bhola Tarr 8.3
11	Jonathan Demme 8.6	Wen Jiang 8.3
12	Luc Besson 8.6	Rakeysh Omprakash Mehra 8.3
13	Roberto Benigni 8.6	Rajkumar Hirani 8.3
14	Bryan Singer 8.6	Lee Unkrich 8.3
15	Hayao Miyazaki 8.6	Vikas Bahl 8.3
16	Anurag Kashyap 8.6	Ashutosh Gowariker 8.25
17	Olivier Nakache 8.6	Pete Docter 8.2
18	James Cameron 8.5	Sujoy Ghosh 8.2
19	Roman Polanski 8.5	S.S. Rajamouli 8.2
20	Ridley Scott 8.5	Quentin Tarantino 8.163
21	Giuseppe Tornatore 8.5	Hayao Miyazaki 8.14
22	Roger Allers 8.5	Andrei Tarkovsky 8.1
23	Tony Kaye 8.5	Kaige Chen 8.1
24	Marco Tullio Giordana 8.5	Milcho Manchevski 8.1

- Есть вопросы

Новый топ-25

	Максимальный рейтинг	Средний рейтинг
0	Frank Darabont 9.3	Christopher Nolan 8.322
1	Christopher Nolan 9.0	Pete Docter 8.2
2	Steven Spielberg 8.9	Quentin Tarantino 8.163
3	Quentin Tarantino 8.9	Hayao Miyazaki 8.14
4	Peter Jackson 8.9	Asghar Farhadi 8.05
5	Robert Zemeckis 8.8	Frank Darabont 7.975
6	David Fincher 8.8	Emir Kusturica 7.933
7	Martin Scorsese 8.7	Krzysztof Kieslowski 7.925
8	Lana Wachowski 8.7	James Cameron 7.883
9	Fernando Meirelles 8.7	Satoshi Kon 7.85
10	Jonathan Demme 8.6	Tom McCarthy 7.833
11	Luc Besson 8.6	Giuseppe Tornatore 7.8
12	Roberto Benigni 8.6	Peter Jackson 7.782
13	Bryan Singer 8.6	Karan Johar 7.767
14	Hayao Miyazaki 8.6	Alejandro González Iñárritu 7.767
15	James Cameron 8.5	John Carney 7.767
16	Roman Polanski 8.5	Ki-duk Kim 7.767
17	Ridley Scott 8.5	Sanjay Leela Bhansali 7.767
18	Giuseppe Tornatore 8.5	David Fincher 7.75
19	Mel Gibson 8.4	Abbas Kiarostami 7.74
20	Jean-Pierre Jeunet 8.4	Claude Berri 7.733
21	Sam Mendes 8.4	Shane Meadows 7.7
22	Chan-wook Park 8.4	Dean DeBlois 7.7
23	Andrew Stanton 8.4	Éric Rohmer 7.683
24	Asghar Farhadi 8.4	Lukas Moodysson 7.667

Жанры и рейтинг

