Analysis of the IMDB movies

by Andras Petre

Introducion

- In this analysis we took the TOP 100 movies according to the IMDB's user ratings
- This project aims to analyze the correlation between the budget and the revenue and discover some statistical values in the different information of a movie (year, ROI (Return Of Investment), director etc.)
- The main findings in the data have been presented in graphical format and the data analysis has been carried out in Python
- The Python libraries, which were used include Pandas, Numpy and BeautifulSoup

The dataset after scraping

	title	year	director	budget	gross_ww	roi
0	The Shawshank Redemption	1994	Frank Darabont	25,000,000.00	28,817,291.00	1.15
1	The Godfather	1972	Francis Ford Coppola	6,000,000.00	246,120,986.00	41.02
2	The Dark Knight	2008	Christopher Nolan	185,000,000.00	1,005,973,645.00	5.44
3	The Godfather: Part II	1974	Francis Ford Coppola	13,000,000.00	48,035,783.00	3.70
4	12 Angry Men	1957	Sidney Lumet	350,000.00	955.00	0.00
...	...	...	...	...	...	...
95	Vertigo	1958	Alfred Hitchcock	2,479,000.00	7,797,728.00	3.15
96	Singin' in the Rain	1952	Stanley Donen	2,540,800.00	1,865,056.00	0.73
97	Bicycle Thieves	1948	Vittorio De Sica	133,000.00	436,655.00	3.28
98	Citizen Kane	1941	Orson Welles	839,727.00	1,603,237.00	1.91
99	M	1931	Fritz Lang	35,566.00	0.00	0.00

- The "title", "year", "director", "budget", "gross_ww" (Gross Worldwide) have been scraped direct from the IMDB webpage

- The calculated ROI (Return Of Investment) shows already here, that there is a low correlation between the ranking (=IMDB rating) and ROI-s

TOP 10 ROI-s

```
#TOP 10 best ROI-s
all_movies_df.sort_values(by='roi', ascending = False).head(10)
```
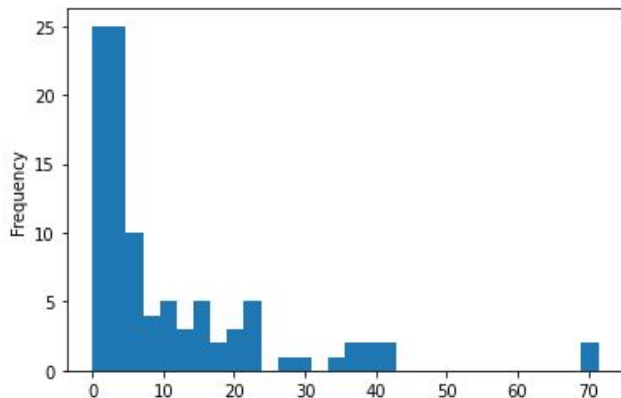
	title	year	director	budget	gross_ww	roi
87	Come and See	1985	Elem Klimov	71,909.00	20,929,068.00	291.05
54	Your Name.	2016	Makoto Shinkai	5,017,246.00	358,180,115.00	71.39
27	Star Wars: Episode IV - A New Hope	1977	George Lucas	11,000,000.00	775,398,007.00	70.49
32	The Intouchables	2011	Olivier Nakache	10,198,820.00	426,588,510.00	41.83
1	The Godfather	1972	Francis Ford Coppola	6,000,000.00	246,120,986.00	41.02
46	Psycho	1960	Alfred Hitchcock	806,947.00	32,041,845.00	39.71
64	The Lives of Others	2006	Florian Henckel von Donnersmarck	2,000,000.00	77,356,942.00	38.68
47	Rear Window	1954	Alfred Hitchcock	1,000,000.00	37,034,514.00	37.03
17	One Flew Over the Cuckoo's Nest	1975	Milos Forman	3,000,000.00	109,118,959.00	36.37
67	Princess Mononoke	1997	Hayao Miyazaki	4,845,631.00	169,785,704.00	35.04

- There is only one movie (The Godfather) from the TOP 10 according to the IMDB rating, which come into the TOP 10 according to the ROI-s

- The best ROI value has "Come and See" which is "only" the 87th best movie according to the users rating

The ROI distribution

```
#ROI distributions
all_movies_df[all_movies_df.roi < 200].roi.plot.hist(bins=30)
```

```
<AxesSubplot:ylabel='Frequency'>
```


- The most movies ROI values are under 10, some movies have an ROI between 30 and 40

- There is a few outlier, which has an ROI about 70 and one movie has over 200, which were removed from the analysis, to avoid bias

TOP 10 budgets

```
#TOP 10 budgets
all_movies_df.sort_values(by='budget', ascending = False).head(10)
```

	title	year	director	budget	gross_ww	roi
56	Avengers: Endgame	2019	Anthony Russo	356,000,000.00	2,797,501,328.00	7.86
57	Avengers: Infinity War	2018	Anthony Russo	321,000,000.00	2,048,359,754.00	6.38
60	The Dark Knight Rises	2012	Christopher Nolan	250,000,000.00	1,081,142,612.00	4.32
2	The Dark Knight	2008	Christopher Nolan	185,000,000.00	1,005,973,645.00	5.44
63	WALL-E	2008	Andrew Stanton	180,000,000.00	521,311,890.00	2.90
58	Coco	2017	Lee Unkrich	175,000,000.00	807,817,888.00	4.62
19	Interstellar	2014	Christopher Nolan	165,000,000.00	701,729,206.00	4.25
8	Inception	2010	Christopher Nolan	160,000,000.00	836,836,967.00	5.23
36	Gladiator	2000	Ridley Scott	103,000,000.00	465,380,802.00	4.52
41	Terminator 2: Judgment Day	1991	James Cameron	102,000,000.00	520,881,154.00	5.11

- There is only 2 movies ("The Dark Knight", "Inception") from the TOP 10 IMDB ratings, which come to the list of the movies with the most budgets

- The TOP 3 movies according to the budgets are "only" between the rankings 50 and 60 in the users ratings
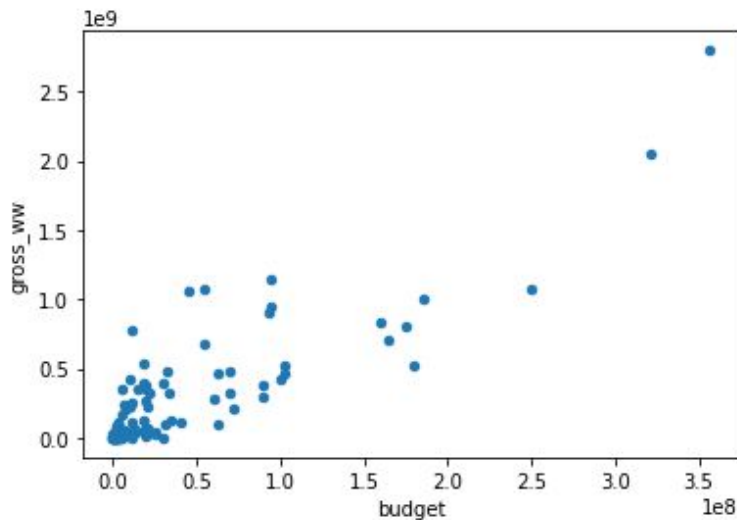
TOP 10 revenue (worldwide)

```
#TOP 10 gross_ww
all_movies_df.sort_values(by='gross_ww', ascending = False).head(10)
```

	title	year	director	budget	gross_ww	roi
56	Avengers: Endgame	2019	Anthony Russo	356,000,000.00	2,797,501,328.00	7.86
57	Avengers: Infinity War	2018	Anthony Russo	321,000,000.00	2,048,359,754.00	6.38
5	The Lord of the Rings: The Return of the King	2003	Peter Jackson	94,000,000.00	1,146,030,912.00	12.19
60	The Dark Knight Rises	2012	Christopher Nolan	250,000,000.00	1,081,142,612.00	4.32
53	Joker	2019	Todd Phillips	55,000,000.00	1,074,419,384.00	19.53
40	The Lion King	1994	Roger Allers	45,000,000.00	1,063,611,805.00	23.64
2	The Dark Knight	2008	Christopher Nolan	185,000,000.00	1,005,973,645.00	5.44
13	The Lord of the Rings: The Two Towers	2002	Peter Jackson	94,000,000.00	947,495,095.00	10.08
10	The Lord of the Rings: The Fellowship of the Ring	2001	Peter Jackson	93,000,000.00	897,690,072.00	9.65
8	Inception	2010	Christopher Nolan	160,000,000.00	836,836,967.00	5.23

- The first two movies are the same as in the TOP 10 budgets (previous slide), so there is a correlation recognizable
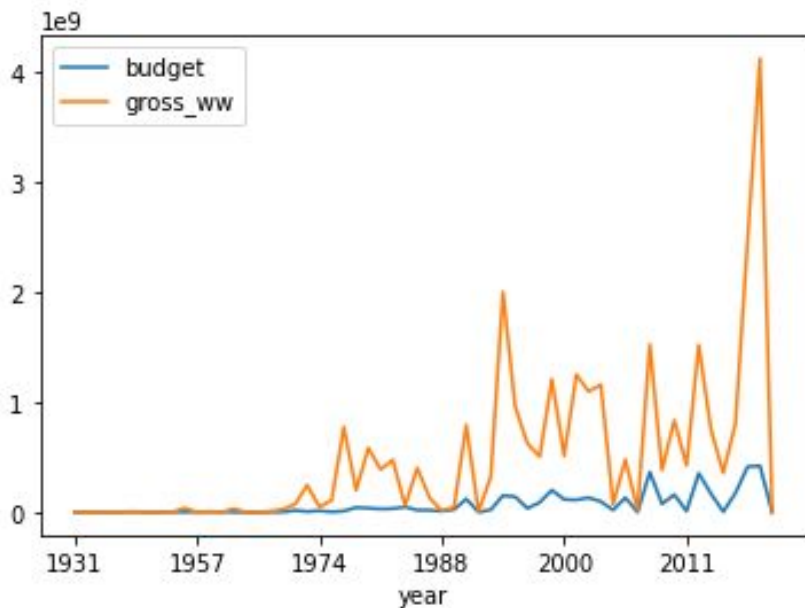
Correlation between budget and revenue

	budget	gross_ww
budget	1.00	0.86
gross_ww	0.86	1.00

- The diagram above shows us that there are some correlation at least on the top of the list (first 30-50 movies)

- The calculated correlation is 86% (below), which is a strong positive relationship between budget and revenue in the TOP 100 movies

Trends

- The diagram shows us an increasing budget and revenue (gross_ww) after the 1960's

- The revenue has a huge growth from the 1990's also in relationship to the budget

- We can discover some exceptional years, when the two values were almost the same

TOP 5 directors

director	title
Christopher Nolan	6
Alfred Hitchcock	4
Quentin Tarantino	4
Stanley Kubrick	4
Steven Spielberg	3

These 5 people have directed 21% of the TOP 100 movies

Conclusion

- We can discover a strong positive (86%) correlation between the given movie's budget and revenue.
- On the other side there are no strong relationship between the movie's ranking and the ROI (Return Of Investment)
- The ROI values are under 10 in the most cases, however there some outlier, which can bias our analysys
- We can see a clear increase of both the budget and the revenue after the 1970's
- All these analysis is related "only" to the TOP 100 movies of IMDB's user ranking

Thank you

Contact details:
andras.petre@gmail.com