

IMDb Top Films Analysis

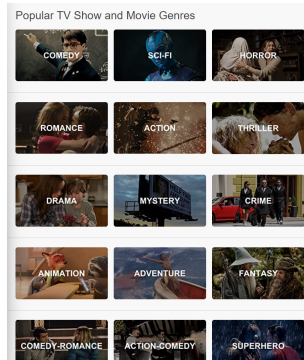
Team Good-2-Great

The University of Texas at Austin

March 2022

Data Collection

- Data source: <https://www.imdb.com>
- Packages used: Beautiful Soup and requests
- Scraped over 1,700 observations from movies by genre
- Excluded observations with incomplete fields



- Removed white space and unnecessary symbols
- Converted variables to desired type
- Created dummy variable columns for each genre

Analysis: Rating vs Genre

- Certain genres have an effect on IMDb score and Metascore
- Several variables were statistically significant
- Results for IMDb score and Metascore were very similar

	coef	P> t
const	64.1331	0.000
4 Action	-5.8758	0.000
Adventure	1.4423	0.267
5 Animation	5.7561	0.001
Biography	1.5397	0.351
3 Comedy	-6.2337	0.000
Crime	-0.6897	0.613
Drama	5.0438	0.000
Family	-4.3629	0.010
Fantasy	-4.4684	0.006
History	1.7306	0.449
1 Horror	-10.6728	0.000
Music	-5.5833	0.008
2 Musical	9.9627	0.009
Mystery	-4.7306	0.003
Romance	-5.4468	0.001
Sci-Fi	0.3418	0.833
Sport	-4.8502	0.073
Thriller	-1.2805	0.380
War	4.2737	0.121

Figure: Regression of Metascore on Genre

	coef	P> t
const	7.2568	0.000
Action	-0.2885	0.000
Adventure	-0.0635	0.352
Animation	0.2121	0.016
Biography	0.1049	0.226
2 Comedy	-0.5478	0.000
Crime	-0.0143	0.842
Drama	0.2167	0.001
3 Family	-0.4554	0.000
Fantasy	-0.1377	0.104
History	-0.0114	0.924
1 Horror	-0.8923	0.000
Music	-0.3862	0.000
4 Musical	0.4464	0.025
Mystery	-0.1651	0.045
5 Romance	-0.3664	0.000
Sci-Fi	-0.0102	0.905
Sport	-0.2258	0.112
Thriller	-0.1711	0.026
War	0.1472	0.309

Figure: Regression of IMDb Score on Genre

Analysis: Revenue vs Rating

- Metascore coefficient is statistically significant
- An increase in Metascore by 1 point leads to about a \$1.29 mil increase in gross revenue, ceteris paribus

	coef	P> t
const	87.9106	0.000
Metascore	1.2909	0.000
Action	-2.0934	0.758
Adventure	33.5627	0.000
Animation	-8.0631	0.373
Biography	-53.8132	0.000
Comedy	-27.5179	0.000
Crime	-28.8019	0.000
Drama	-33.9594	0.000
Family	-24.1156	0.008
Fantasy	27.0397	0.002
History	-38.4057	0.002
Horror	-61.6635	0.000
Music	-51.8105	0.000
Musical	-5.4903	0.787
Mystery	-23.2919	0.006
Romance	-5.7183	0.525
Sci-Fi	21.3803	0.014
Sport	-13.8978	0.338
Thriller	-21.8428	0.005
War	-68.5367	0.000

Figure: Regression of Gross Revenue on Metascore

Limitations: Time Constraints

- Domestic gross revenue is not inflation-adjusted
- Potential endogeneity as a result

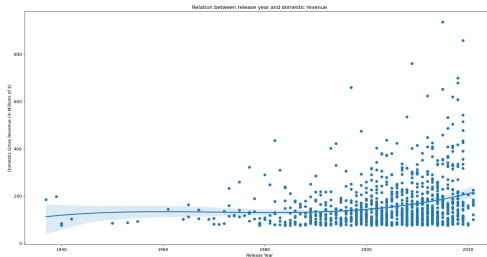


Figure: 1

Limitations: Data Constraints

- Some fields were empty and those observations were dropped
- IMDb not used as much in the distant past (Figure: 2)

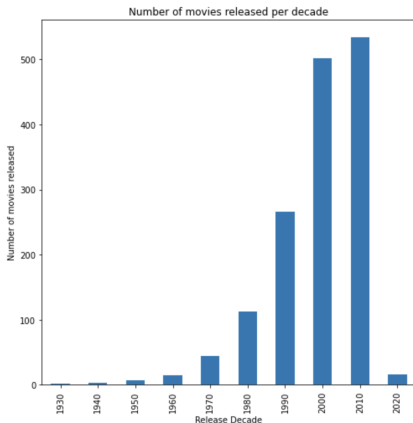


Figure: 2