IS 489-1001

Movie Busts

Jeremy Smoljan

Professor Sutirtha Chatterjee

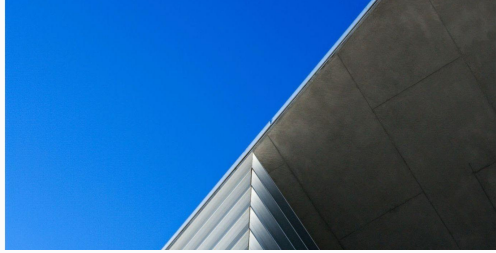
12/12/2025

Contents

A Movie Industry ‘Crisis’ Analysis

Determining specific variables to improve box office gross revenue in the movie industry.

Cleaning and Transforming Data



Utilizing RStudio, Excel, and Rapidminer by Altair's AI Studio

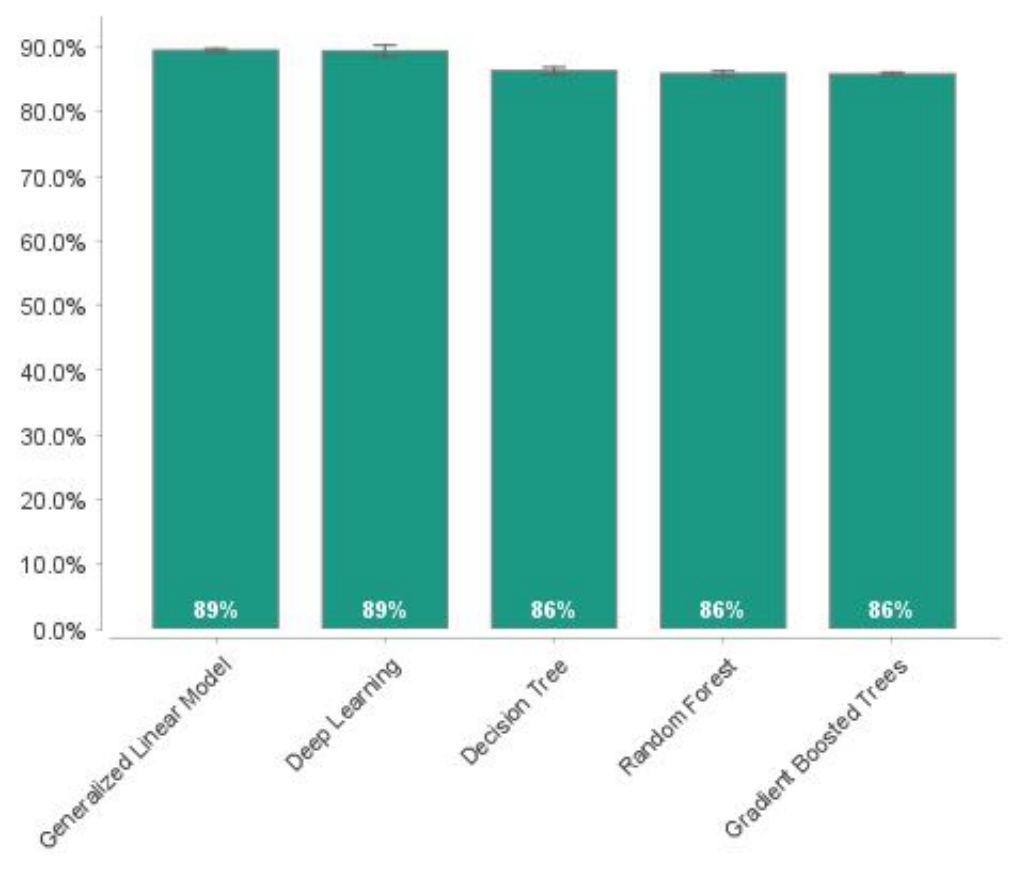
Inconsistent Data

Cleaning the data was an important factor in appropriately studying and analyzing variable importance.

Data Formatting

Reformatting data columns into directly analyzable formats was vital.

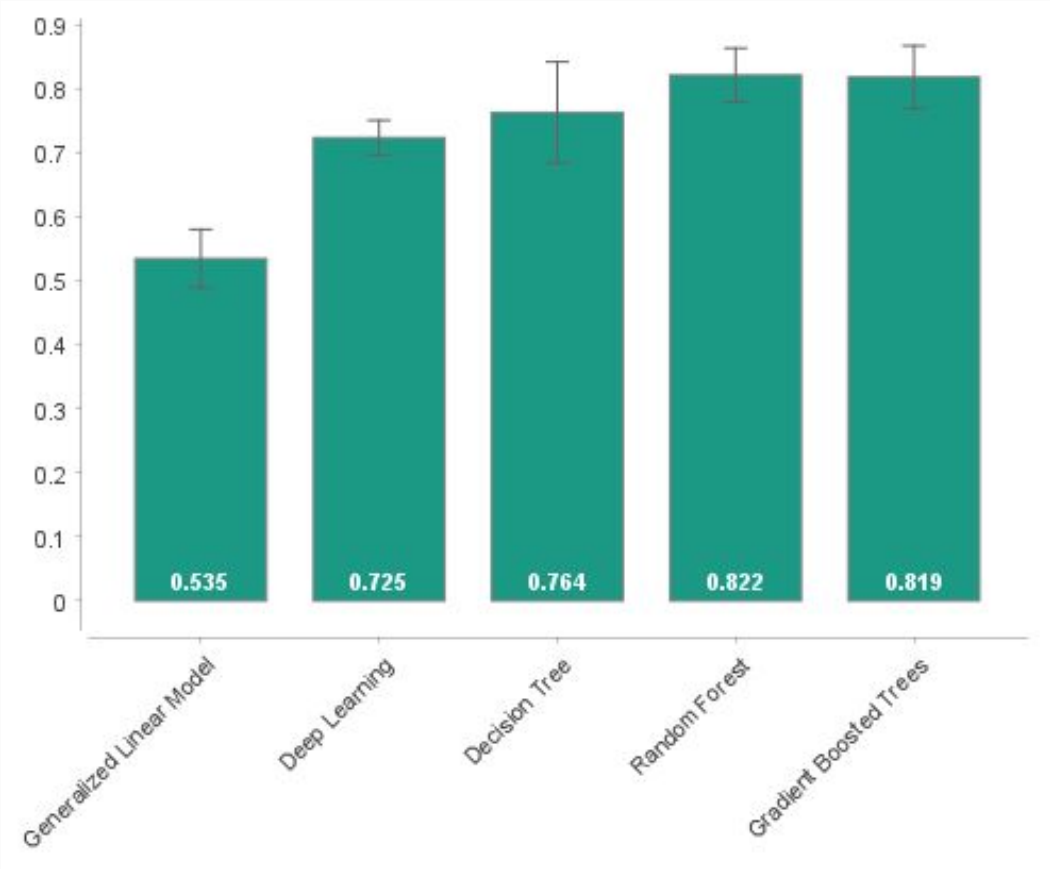
Model Relative Error Rates



Traits included:

- Budget
- Metacritic Rating
- ‘Age’ Rating
- Season - Yearly
- Remake
- IMDB Rating
- Release Month
- Sequel
- Tent-Pole Production

Model Correlation



Representing the overall correlation between a movies box office gross revenue and the independent variables provided.

Independent Variable Importance

Gradient Boosted Trees - Weights

Attribute	Weight
Budget	0.415
metacritic	0.249
rating	0.083
season	0.031
remake	0.024
imdb_rating	0.004
release_month	0.000
sequel	0.000
tentpole	0

Weights of specified independent variables found within the Gradient Boosted Tree Model

root_mean_squared_error

```
root_mean_squared_error: 26729406.044 +/- 2292388.804
```

correlation

```
correlation: 0.819 +/- 0.049 (micro average: 0.818)
```

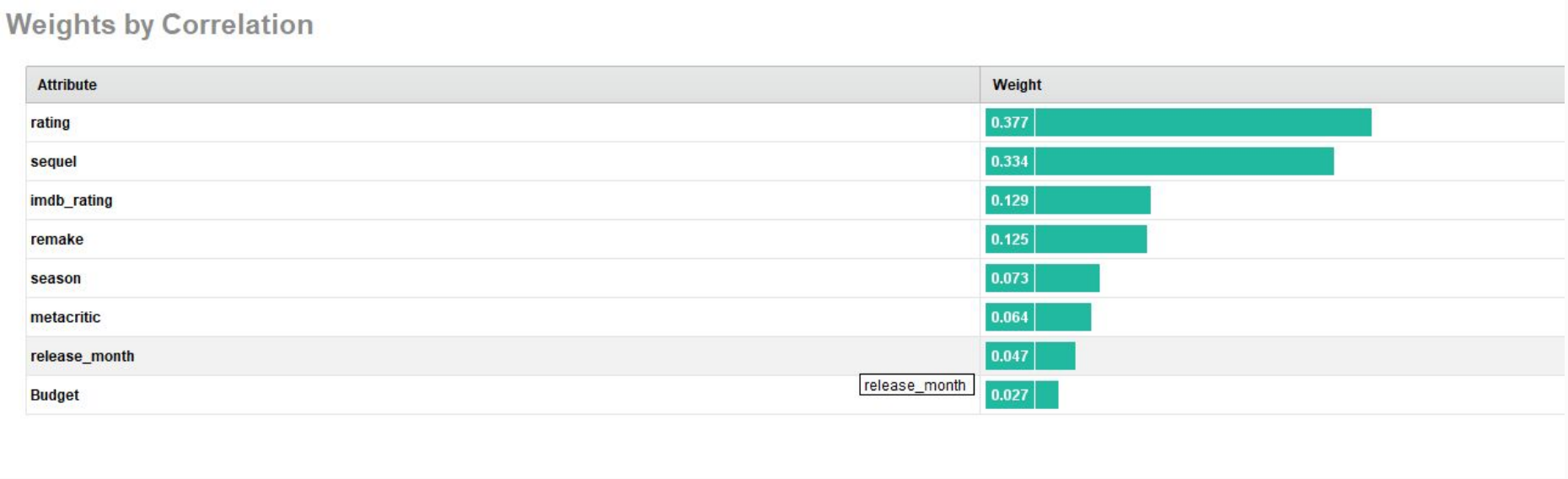
Independent Variable Importance

Random Forest - Weights

Attribute	Weight
Budget	0.614
metacritic	0.320
remake	0.104
rating	0.033
season	0.031
imdb_rating	0.019
release_month	0.003
sequel	0.000
tentpole	0

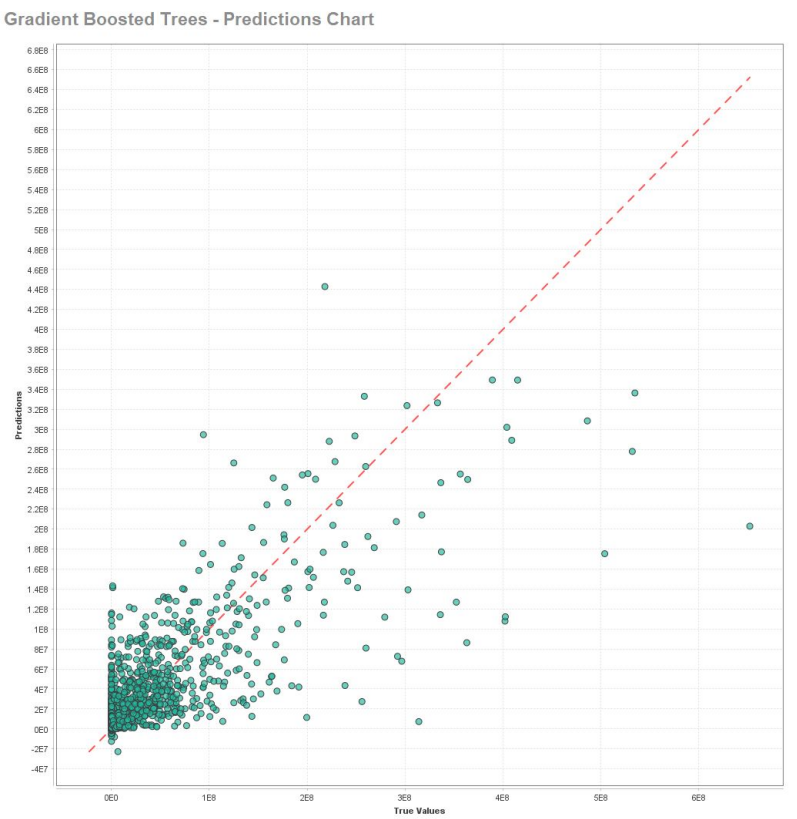
Weights of specified independent variables found within the Random Forest Regression Model

Combined Model Variable Importance



Highlighting a 5 separate model combination of weights for independent variables

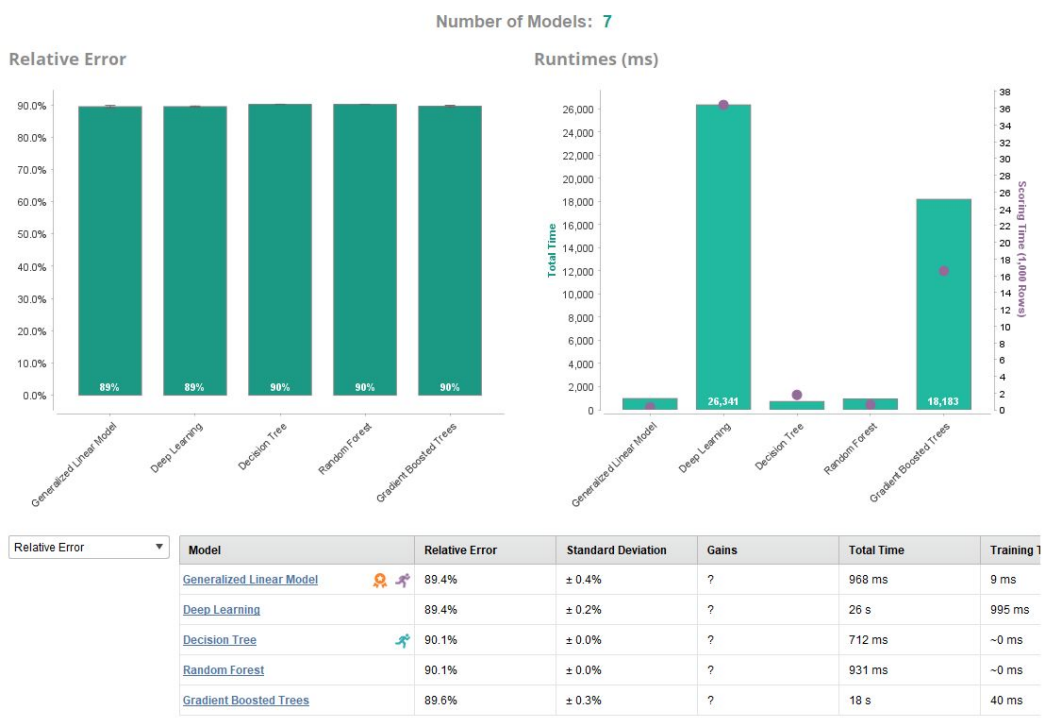
Predictions and Variable Correlations



Correlations

Attributes ↑	Box...	Bud...	imd...	me...	ratin...	ratl...	ratin...	ratin...	rating = ...	rating = ...	rating = ...	rating = ...	rating = ...	rating = R	rating = ...	rating = ...	rating = ...	rating = ...	rating = ...	rating = ...	rating = ...	rating = X	release...	remake	season ...	season ...	season ...	seq...
Box.Office.Gross	1	0.027	0.129	0.064	0.010	0.077	-0.006	-0.237	-0.009	-0.089	-0.013	0.168	0.281	-0.009	-0.017	-0.006	-0.020	-0.010	-0.003	-0.003	-0.045	0.000	0.047	0.125	-0.058	-0.009	0.060	0.334
Budget	0.027	1	0.012	-0.0...	-0.001	-0.000	-0.001	0.007	-0.002	0.003	-0.001	0.000	0.001	-0.008	-0.003	-0.001	-0.003	-0.002	-0.001	-0.001	-0.003	-0.001	-0.013	0.007	-0.005	0.007	0.007	0.010
imdb_rating	0.129	0.012	1	0.470	0.022	0.023	-0.000	0.046	0.002	0.016	0.019	-0.018	-0.020	-0.048	-0.005	0.006	-0.012	0.003	-0.001	-0.000	0.027	-0.003	0.016	-0.043	0.032	0.006	-0.028	-0.046
metacritic	0.064	-0.006	0.470	1	0.000	0.037	0.000	0.088	-0.006	0.074	0.031	-0.032	-0.127	-0.059	0.018	-0.011	0.014	0.009	0.000	0.000	0.076	0.000	0.015	-0.070	0.015	0.004	-0.002	-0.055
rating = APPROVED	0.010	-0.001	0.022	0.000	1	-0.002	-0.000	-0.015	-0.001	-0.006	-0.001	-0.005	-0.008	-0.011	-0.001	-0.000	-0.001	-0.001	-0.000	-0.000	-0.003	-0.000	-0.011	-0.003	0.002	0.003	-0.010	-0.004
rating = G	0.077	-0.000	0.023	0.037	-0.002	1	-0.002	-0.077	-0.003	-0.029	-0.004	-0.028	-0.043	-0.059	-0.005	-0.002	-0.006	-0.003	-0.001	-0.001	-0.015	-0.001	0.002	-0.017	-0.006	0.006	0.004	0.049
rating = MISSING	-0.006	-0.001	-0.000	0.000	-0.000	-0.002	1	-0.015	-0.001	-0.006	-0.001	-0.005	-0.008	-0.011	-0.001	-0.000	-0.001	-0.001	-0.000	-0.000	-0.003	-0.000	-0.034	-0.003	0.030	-0.011	-0.010	-0.004
rating = NA	-0.237	0.007	0.046	0.088	-0.015	-0.077	-0.015	1	-0.024	-0.232	-0.035	-0.218	-0.339	-0.469	-0.043	-0.015	-0.059	-0.026	-0.009	-0.009	-0.117	-0.009	-0.042	-0.084	0.031	0.006	-0.019	-0.105
rating = NC-17	-0.009	-0.002	0.002	-0.0...	-0.001	-0.003	-0.001	-0.024	1	-0.009	-0.001	-0.009	-0.013	-0.018	-0.002	-0.001	-0.002	-0.001	-0.000	-0.000	-0.005	-0.000	-0.003	-0.005	0.014	-0.010	-0.017	0.010
rating = NOT RATED	-0.089	0.003	0.016	0.074	-0.006	-0.029	-0.006	-0.232	-0.009	1	-0.013	-0.083	-0.129	-0.178	-0.016	-0.006	-0.019	-0.010	-0.003	-0.003	-0.044	-0.003	-0.009	-0.011	-0.001	0.012	0.000	-0.039
rating = NR	-0.013	-0.001	0.019	0.031	-0.001	-0.004	-0.001	-0.035	-0.001	-0.013	1	-0.013	-0.019	-0.027	-0.002	-0.001	-0.003	-0.001	-0.000	-0.000	-0.007	-0.000	0.004	-0.008	0.001	0.009	0.007	-0.010
rating = PG	0.168	0.000	-0.018	-0.0...	-0.005	-0.028	-0.005	-0.218	-0.009	-0.063	-0.013	1	-0.121	-0.168	-0.016	-0.005	-0.018	-0.009	-0.003	-0.003	-0.042	-0.003	0.040	0.016	-0.006	-0.003	0.011	0.070
rating = PG-13	0.281	0.001	-0.020	-0.1...	-0.008	-0.043	-0.008	-0.339	-0.013	-0.129	-0.019	-0.121	1	-0.260	-0.024	-0.008	-0.028	-0.014	-0.005	-0.005	-0.065	-0.005	0.014	0.089	-0.020	0.001	0.010	0.097
rating = R	-0.009	-0.008	-0.048	-0.0...	-0.011	-0.059	-0.011	-0.469	-0.018	-0.178	-0.027	-0.168	-0.260	1	-0.033	-0.011	-0.039	-0.020	-0.007	-0.007	-0.090	-0.007	0.015	0.034	-0.010	-0.015	0.008	0.017
rating = TV-14	-0.017	-0.003	-0.005	0.018	-0.001	-0.005	-0.001	-0.043	-0.002	-0.016	-0.002	-0.016	-0.024	-0.033	1	-0.001	-0.004	-0.002	-0.001	-0.001	-0.008	-0.001	0.003	-0.010	-0.007	0.005	0.005	-0.003
rating = TV-G	-0.006	-0.001	0.006	-0.0...	-0.000	-0.002	-0.000	-0.015	-0.001	-0.006	-0.001	-0.005	-0.008	-0.011	-0.001	1	-0.001	-0.001	-0.000	-0.000	-0.003	-0.000	0.009	-0.003	0.002	-0.011	0.019	-0.004
rating = TV-MA	-0.020	-0.003	-0.012	0.014	-0.001	-0.006	-0.001	-0.050	-0.002	-0.019	-0.003	-0.018	-0.028	-0.039	-0.004	-0.001	1	-0.002	-0.001	-0.001	-0.010	-0.001	-0.007	-0.011	-0.008	0.023	-0.005	-0.015
rating = TV-PG	-0.010	-0.002	0.003	0.009	-0.001	-0.003	-0.001	-0.026	-0.001	-0.010	-0.001	-0.009	-0.014	-0.020	-0.002	-0.001	-0.002	1	-0.000	-0.000	-0.005	-0.000	-0.001	-0.006	0.003	-0.003	-0.009	-0.008
rating = TV-Y	-0.003	-0.001	-0.001	0.000	-0.000	-0.001	-0.000	-0.009	-0.000	-0.003	-0.000	-0.003	-0.005	-0.007	-0.001	-0.000	-0.001	-0.000	1	-0.000	-0.002	-0.000	0.005	-0.002	-0.007	-0.007	0.020	-0.003
rating = UNRATED	-0.045	-0.003	0.027	0.076	-0.003	-0.015	-0.003	-0.117	-0.005	-0.044	-0.007	-0.042	-0.065	-0.090	-0.008	-0.003	-0.010	-0.005	-0.002	-0.002	1	-0.002	0.008	-0.026	-0.003	-0.005	-0.006	-0.020
rating = Unrated	-0.003	-0.001	-0.000	0.000	-0.000	-0.001	-0.000	-0.009	-0.000	-0.003	-0.000	-0.003	-0.005	-0.007	-0.001	-0.000	-0.001	-0.000	-0.000	1	-0.002	-0.000	-0.004	-0.002	-0.007	0.018	-0.006	-0.003
rating = X	0.000	-0.001	-0.003	0.000	-0.000	-0.001	-0.000	-0.009	-0.000	-0.003	-0.000	-0.003	-0.005	-0.007	-0.001	-0.000	-0.001	-0.000	-0.000	-0.000	-0.002	1	-0.014	-0.002	-0.007	-0.007	-0.006	0.047
release_month	0.047	-0.013	0.016	0.015	-0.011	0.002	-0.034	-0.042	-0.003	-0.009	0.004	0.040	0.014	0.015	0.003	0.009	-0.007	-0.001	0.005	-0.004	0.008	-0.014	1	0.013	0.526	-0.406	0.113	0.037
remake	0.125	0.007	-0.043	-0.0...	-0.003	-0.017	-0.003	-0.084	-0.005	-0.011	-0.008	0.016	0.089	0.034	-0.010	-0.003	-0.011	-0.006	-0.002	-0.002	-0.026	-0.002	0.013	1	-0.017	-0.004	0.002	0.039
season = Fall	-0.058	-0.005	0.032	0.015	0.002	-0.006	0.030	0.031	0.014	-0.001	0.001	-0.006	-0.020	-0.010	-0.007	0.002	-0.008	0.003	-0.007	-0.007	-0.003	-0.007	0.526	-0.017	1	-0.385	-0.349	-0.014
season = Spring	-0.009	0.007	0.006	0.004	0.003	0.006	-0.011	0.006	-0.010	0.012	0.009	-0.003	0.001	-0.015	0.005	-0.011	0.023	-0.003	-0.007	0.018	-0.005	-0.007	-0.406	-0.004	-0.385	1	-0.332	-0.013
season = Summer	0.060	0.007	-0.028	-0.0...	-0.010	0.004	-0.010	-0.019	-0.017	0.000	0.007	0.011	0.010	0.008	0.005	0.019	-0.005	-0.009	0.020	-0.006	-0.006	-0.006	0.113	0.002	-0.349	-0.332	1	0.041
sequel	0.334	0.010	-0.046	-0.0...	-0.004	0.049	-0.004	-0.105	0.010	-0.039	-0.010	0.070	0.097	0.017	-0.003	-0.004	-0.015	-0.008	-0.003	-0.003	-0.020	0.047	0.037	0.039	-0.014	-0.013	0.041	1

Genre to Revenue Modeling



Relative Error

Model	Relative Error	Standard Deviation	Gains	Total Time	Training T
Generalized Linear Model	89.4%	± 0.4%	?	968 ms	9 ms
Deep Learning	89.4%	± 0.2%	?	26 s	995 ms
Decision Tree	90.1%	± 0.0%	?	712 ms	~0 ms
Random Forest	90.1%	± 0.0%	?	931 ms	~0 ms
Gradient Boosted Trees	89.6%	± 0.3%	?	18 s	40 ms

Highlighting the Relative Error rate between genre category to revenue predictions.

Genre Predictions to Gross Revenue

Gradient Boosted Trees - Predictions

Row No.	Box.Office.Gross	prediction(Box.Office.Gross)	genre
1993	652270625	80511462.093	Action, Adventure, Sci-Fi
1721	534858444	13037318.994	Action, Crime, Drama
267	532177324	80511462.093	Action, Adventure, Sci-Fi
2912	504014165	6161822.339	MISSING
1992	486295561	62498287.249	Animation, Adventure, Comedy
1991	415004880	62498287.249	Animation, Adventure, Comedy
2291	409013994	80511462.093	Action, Adventure, Sci-Fi
1990	404000714	72073646.843	Action, Adventure, Fantasy
1989	402453882	131685325.311	Adventure, Sci-Fi, Thriller
1988	402111870	80511462.093	Action, Adventure, Sci-Fi
2290	389213281	80511462.093	Action, Adventure, Sci-Fi
2631	364001123	32267421.999	Adventure, Drama, Family
3124	363070709	23135077.760	Action, Adventure, Comedy
1987	356461711	62498287.249	Animation, Adventure, Comedy
1986	352390543	80511462.093	Action, Adventure, Sci-Fi
538	337135885	80511462.093	Action, Adventure, Sci-Fi
2289	336530303	38183549.374	Action, Adventure
1720	336045770	51006776.624	Animation, Action, Adventure
1470	333176600	80511462.093	Action, Adventure, Sci-Fi
2288	317101119	72073646.843	Action, Adventure, Fantasy
1719	314057748	80511462.093	Action, Adventure, Sci-Fi
266	303003568	78394869.374	Adventure, Fantasy
1718	301959197	61055269.405	Adventure, Family, Fantasy
537	296623634	123249466.593	Adventure, Drama, Fantasy
536	292298923	123249466.593	Adventure, Drama, Fantasy
1985	291045518	72073646.843	Action, Adventure, Fantasy
265	279261160	17600871.174	Comedy, Romance
1984	268492764	62498287.249	Animation, Adventure, Comedy
1717	262030663	38183549.374	Action, Adventure
535	260044825	24346662.393	Comedy, Family, Fantasy
2630	259766572	80511462.093	Action, Adventure, Sci-Fi
264	258366855	78394869.374	Adventure, Fantasy
534	255959475	14127361.582	Biography, Drama, Sport

Gradient Boosted Trees - Predictions

Row No.	Box.Office.Gross	prediction(Box.Office.Gro...	genre
1466	113203870	131685325.311	Action, Adventure, Mystery
1989	402453882	131685325.311	Adventure, Sci-Fi, Thriller
2627	89021735	131685325.311	Action, Adventure, Mystery
2783	0	131685325.311	Adventure, Sci-Fi, Thriller
2898	66002193	131685325.311	Action, Adventure, Mystery
523	124987023	123249466.593	Adventure, Drama, Fantasy
536	292298923	123249466.593	Adventure, Drama, Fantasy
537	296623634	123249466.593	Adventure, Drama, Fantasy
789	2184640	123249466.593	Adventure, Drama, Fantasy
526	184208848	121922920.561	Adventure, Drama, Western
2378	0	121922920.561	Adventure, Drama, Western
870	228433663	84160296.030	Adventure, Drama, Sci-Fi
76	0	80511462.093	Action, Adventure, Sci-Fi
267	532177324	80511462.093	Action, Adventure, Sci-Fi
381	0	80511462.093	Action, Adventure, Sci-Fi
538	337135885	80511462.093	Action, Adventure, Sci-Fi
761	6820	80511462.093	Action, Adventure, Sci-Fi
1470	333176600	80511462.093	Action, Adventure, Sci-Fi
1705	132556852	80511462.093	Action, Adventure, Sci-Fi
1708	158848340	80511462.093	Action, Adventure, Sci-Fi
1710	176654505	80511462.093	Action, Adventure, Sci-Fi
1719	314057748	80511462.093	Action, Adventure, Sci-Fi
1825	0	80511462.093	Action, Adventure, Sci-Fi
1843	0	80511462.093	Action, Adventure, Sci-Fi
1953	57761012	80511462.093	Action, Adventure, Sci-Fi
1965	103144286	80511462.093	Action, Adventure, Sci-Fi
1967	116601172	80511462.093	Action, Adventure, Sci-Fi
1972	146408305	80511462.093	Action, Adventure, Sci-Fi
1983	245439076	80511462.093	Action, Adventure, Sci-Fi
1986	352390543	80511462.093	Action, Adventure, Sci-Fi
1988	402111870	80511462.093	Action, Adventure, Sci-Fi
1993	652270625	80511462.093	Action, Adventure, Sci-Fi
2276	155442489	80511462.093	Action, Adventure, Sci-Fi

Difference between Genre Categories found in Actual Gross income (LEFT) and Predicted Gross income (RIGHT).

Genre Matrix Correlation

Correlations

Attributes	Box.Offi...
Box.Office.Gross	1
genre = Action	-0.028
genre = Action, Adventure	0.059
genre = Action, Adventure, Biography	-0.002
genre = Action, Adventure, Comedy	0.040
genre = Action, Adventure, Crime	0.050
genre = Action, Adventure, Drama	0.065
genre = Action, Adventure, Family	0.073
genre = Action, Adventure, Fantasy	0.202
genre = Action, Adventure, History	-0.006
genre = Action, Adventure, Horror	0.024
genre = Action, Adventure, Mystery	0.087
genre = Action, Adventure, Romance	0.002
genre = Action, Adventure, Sci-Fi	0.259
genre = Action, Adventure, Thriller	0.054
genre = Action, Adventure, Western	0.012
genre = Action, Biography, Crime	-0.010
genre = Action, Biography, Drama	0.065
genre = Action, Biography, History	?
genre = Action, Comedy	0.018
genre = Action, Comedy, Crime	0.032
genre = Action, Comedy, Drama	-0.021
genre = Action, Comedy, Family	0.001
genre = Action, Comedy, Fantasy	0.015
genre = Action, Comedy, History	-0.004
genre = Action, Comedy, Horror	-0.006
genre = Action, Comedy, Music	0.003
genre = Action, Comedy, Musical	-0.004

Recommendations

Focus on the Following Factors:

- PG-13 Rated Movies
- PG Rated Movies
- Summer/Winter Releases
- Budget Prioritization
- Action/Adventure Genres
- Sequels then Remakes

AVOID: 'TV' ratings for revenue growth, good weather release dates i.e. Fall/Spring, genres that exclude Action and/or Adventure

03

Will it be a box office bust or a hit?

01 Prioritize dependent variables that are known to increase revenue and improve success.

02 Consider cultural, economic, and political factors when developing movies.

03 Properly budget between marketing, talent, and directors. Known talent and directors can be considered partial marketing.