

Big Box Office Movies

What Makes the Best Movie?

Goals

- To determine the key characteristics of successful movies
- Analyzing the last 100 years worth of movie data to understand which the best genre, best directors, and best runtime
- We will use datasets from IMDB, Box Office Mojo, Rotten Tomatoes, TheMovieDB, and The Numbers in our work

What is 'Best?'

- Overall Profit will be the best indicator of success
- Total Gross
- ROI

Data Exploration

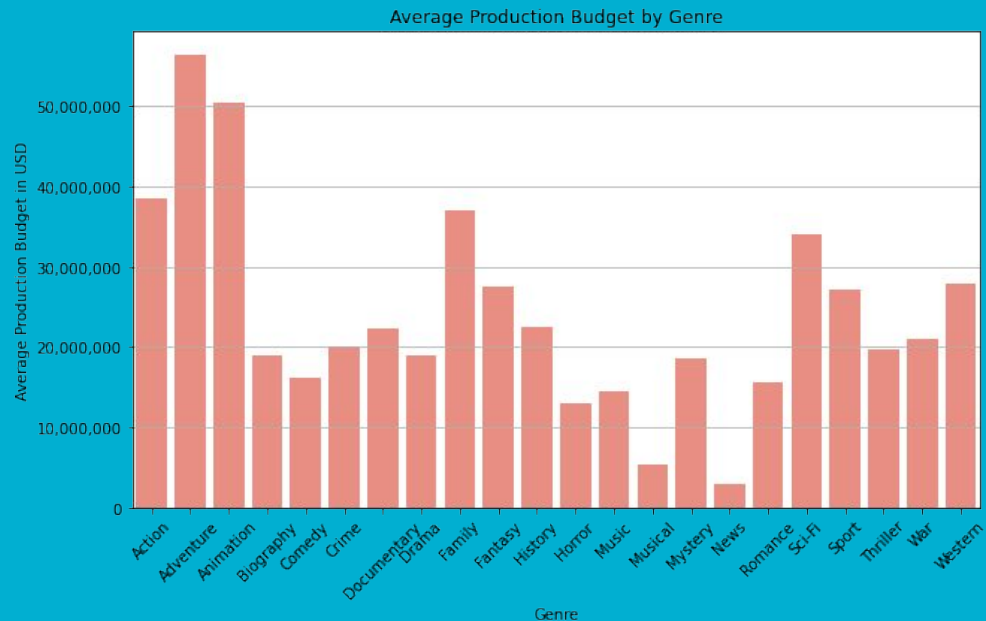
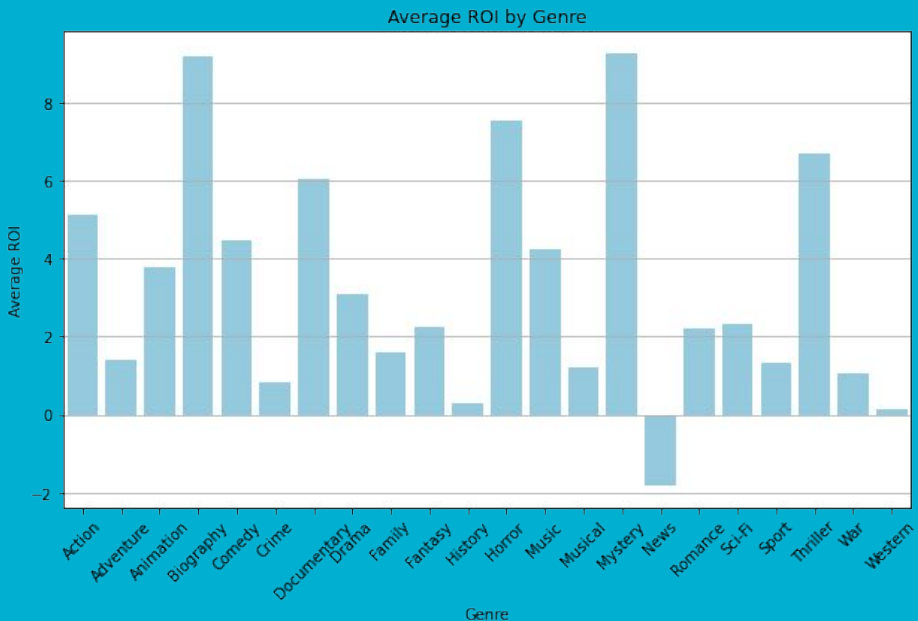
- Clean data 294,000 -> 5,000
- Connecting datasets using SQL JOIN Function
- Standardize columns
 - Dates, directors, runtime

Analysis Methods

- Create graphs to identify trends
- Regression Analysis to predict future trends
- R-squared analysis



Analysis by Genre

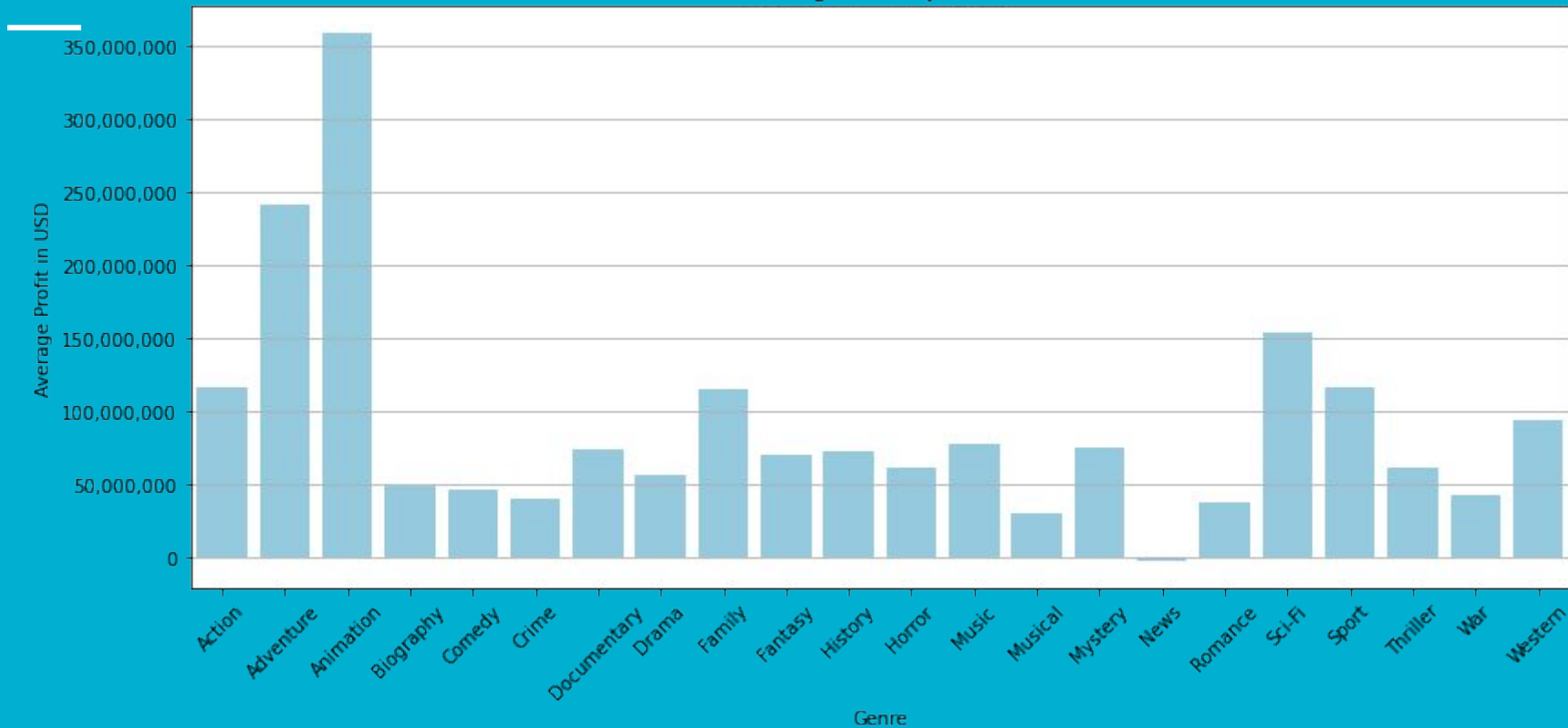


Here we can see the Average ROI (left) and Production Budget (right) per Genre.

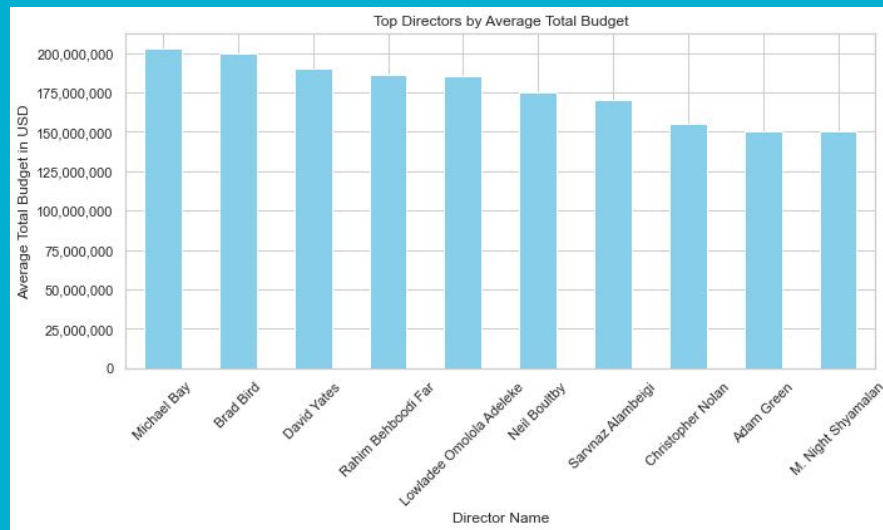
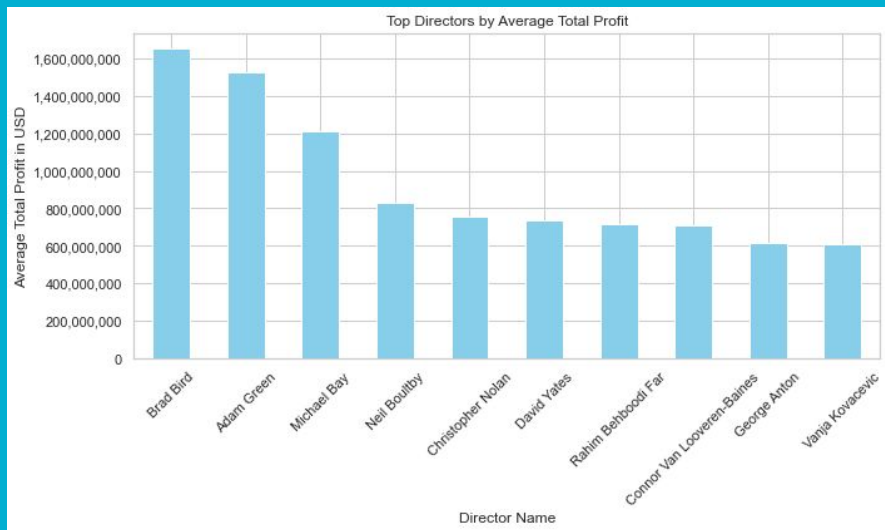
Analysis by Genre

—Notice how higher budgets have lead to higher Profits!

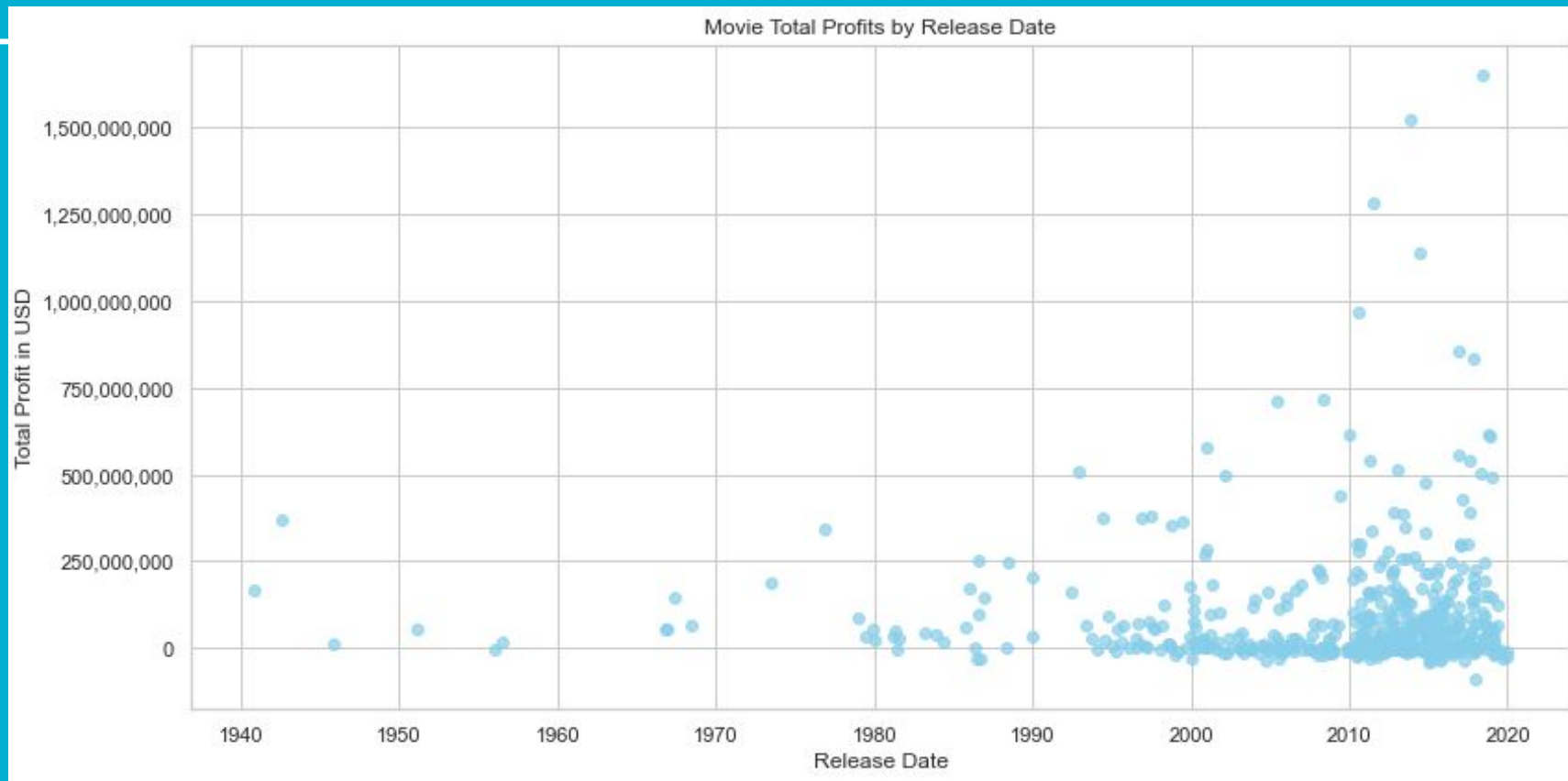
Average Profit by Genre



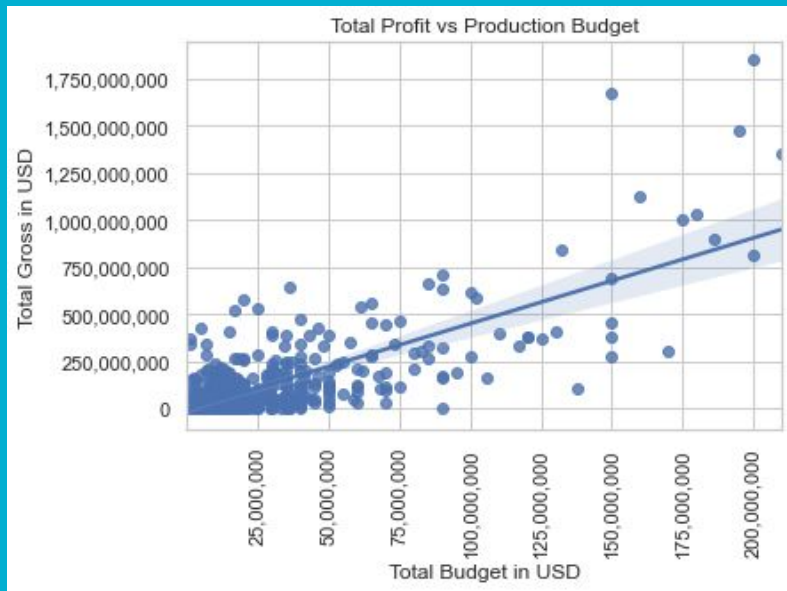
Analysis by Director–6 Directors appear in both charts



Analysis by Release Date—Movies are becoming more profitable



Budget vs Gross



R-Squared of .6 means we can assume 60% of the gross is impacted by the budget

Dep. Variable: total_gross

R-squared: 0.609

Model: OLS

Adj. R-squared: 0.609

Method: Least Squares

F-statistic: 1084.

Date: Mon, 18 Dec 2023

Prob (F-statistic): 5.68e-144

Time: 18:49:11

Log-Likelihood: -12699.

No. Observations: 697

AIC: 2.540e+04

Df Residuals: 695

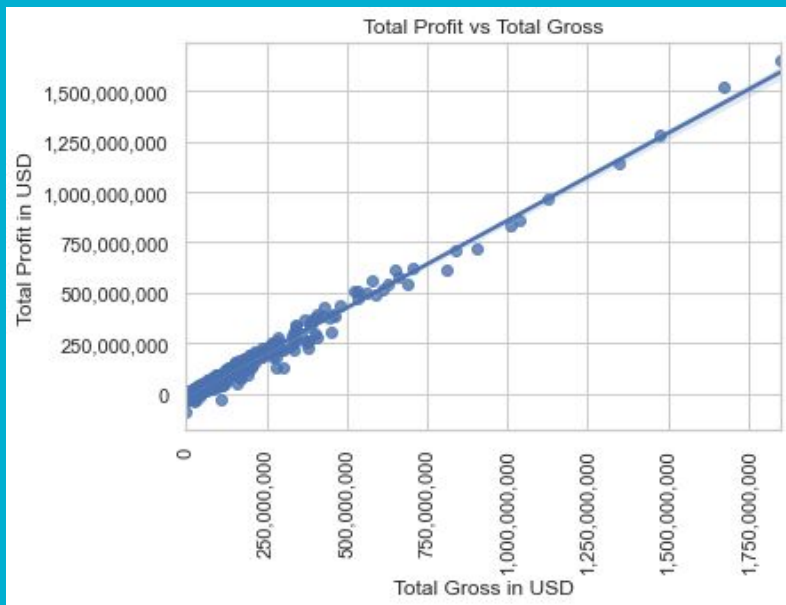
BIC: 2.541e+04

Df Model: 1

Covariance Type: nonrobust

	coef	std err	t	P> t	[0.025	0.975]
Intercept	8.876e+06	8.37e+05	10.605	0.000	7.23e+06	1.05e+07
total_gross	0.1338	0.004	32.922	0.000	0.126	0.142

Gross Vs Profit



Almost a direct 1:1 correlation! The profit is directly impacted by the budget!

Dep. Variable: total_profit

R-squared: 0.985

Model: OLS

Adj. R-squared: 0.985

Method: Least Squares

F-statistic: 4.539e+04

Date: Mon, 18 Dec 2023

Prob (F-statistic): 0.00

Time: 18:49:13

Log-Likelihood: -12699.

No. Observations: 697

AIC: 2.540e+04

Df Residuals: 695

BIC: 2.541e+04

Df Model: 1

Covariance Type: nonrobust

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-8.876e+06	8.37e+05	-10.605	0.000	-1.05e+07	-7.23+06
total_gross	0.8662	0.004	213.059	0.000	0.858	0.874

Limitations

- Lots of incomplete data, so imputation had to be done
- Data sets containing ratings and reviews was not able to be joined
- Largely US Movies

Conclusions

- Budget is the best indicator of success
 - Larger budgets lead to higher gross and more profits
- Successful Directors:
 - Michael Bay
 - Brad Bird
 - David Yates
 - Rahim Behboodi
 - Neil Boulton
 - Adam Green
- Best Genres:
 - Animation (Costly)
 - Sci-Fi (Costly)
 - Mystery (Biggest ROI)

Next Steps

- Compile data on all movies ever made
- Explore timelines of revenue generation for older movies/movies with high viewer ratings/reviews

Thank You For Your Time

- Additional Questions?
- <https://www.linkedin.com/in/will-byrd-40323863>