

Making a (successful) Movie

Analysis of Movie Data

Henry Graham Costello | March 2022

Summary Findings

1. Doesn't have to be good!
2. Sci-Fi, Animation, or Adventure
(or all three!)
3. SUMMER!

The Data

Sources :

Two IMDB Data Sets

Box Office Mojo Gross Revenue Data Set

The MovieDB Data Set

Examining the Data

IMDB Basics Data Set:

146,144 movies

IMDB Ratings Data Set:

73,856 movies

Box Office Mojo Data Set:

3,387 movies

The Movie Database Data Set:

26,517 movies

Cleaning the Data

Rating and Genre graphs:

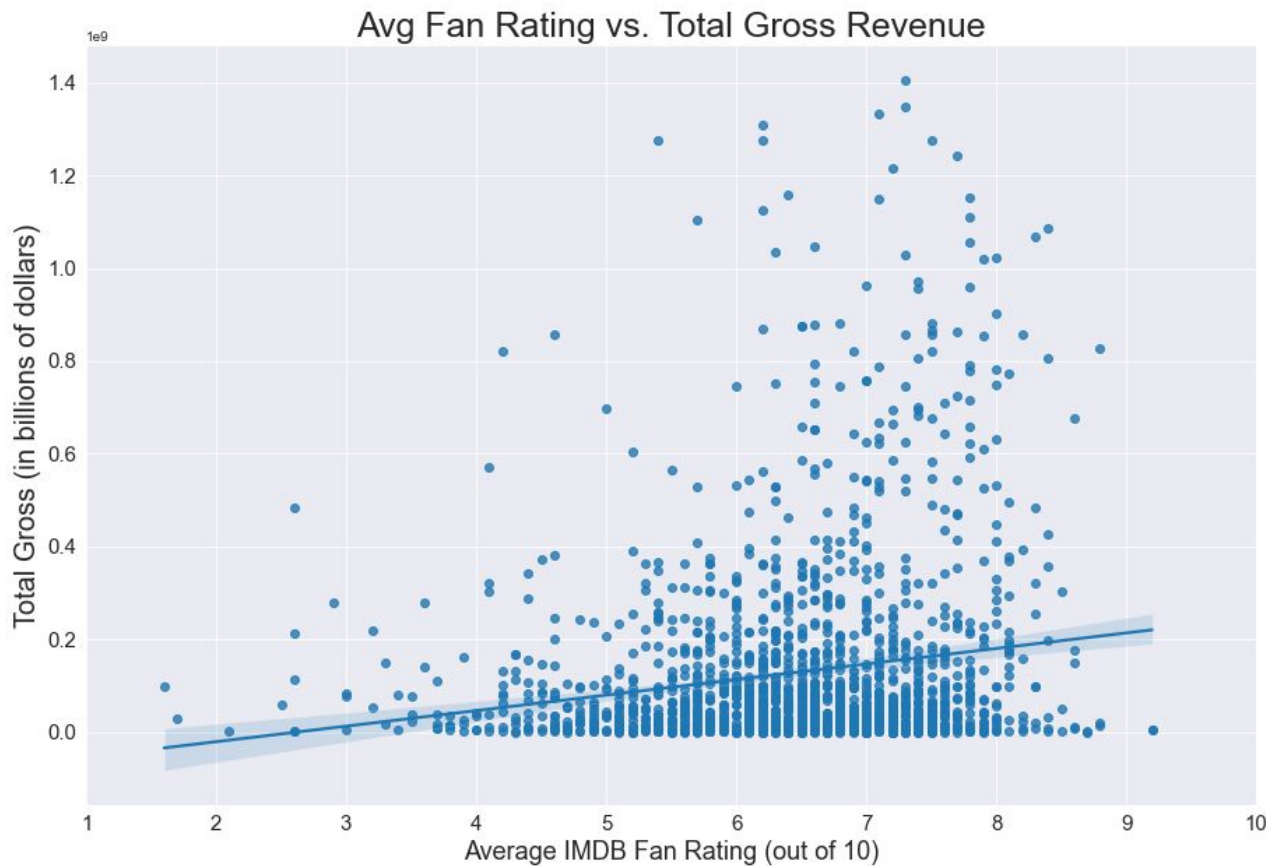
IMDB Data Sets + Box Office Mojo:

1,799 movies

Season graph:

The Movie Database + Box Office Mojo:

1,702 movies

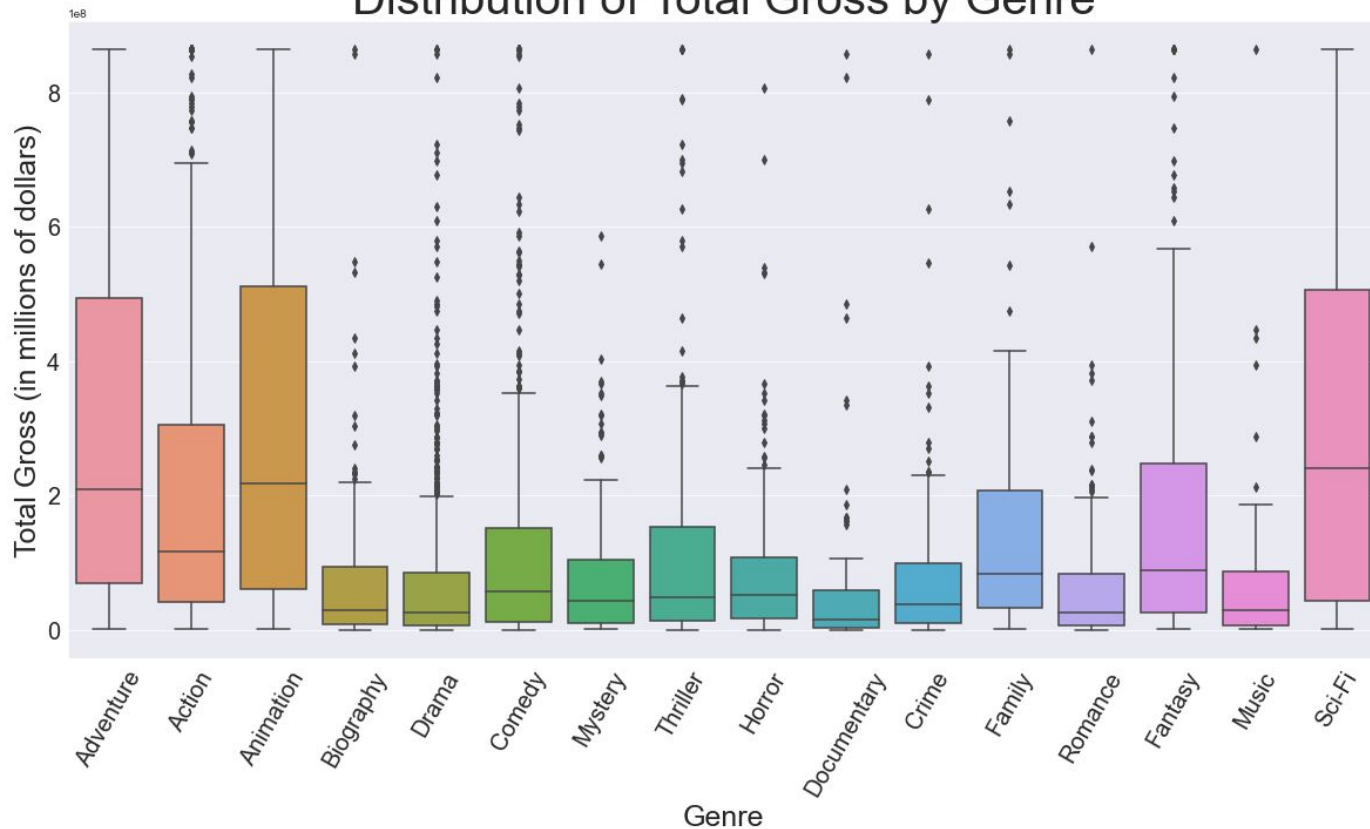


Correlation (r^2)
0.026

Key Takeaways

1. Don't worry about quality
2. Cluster at low gross

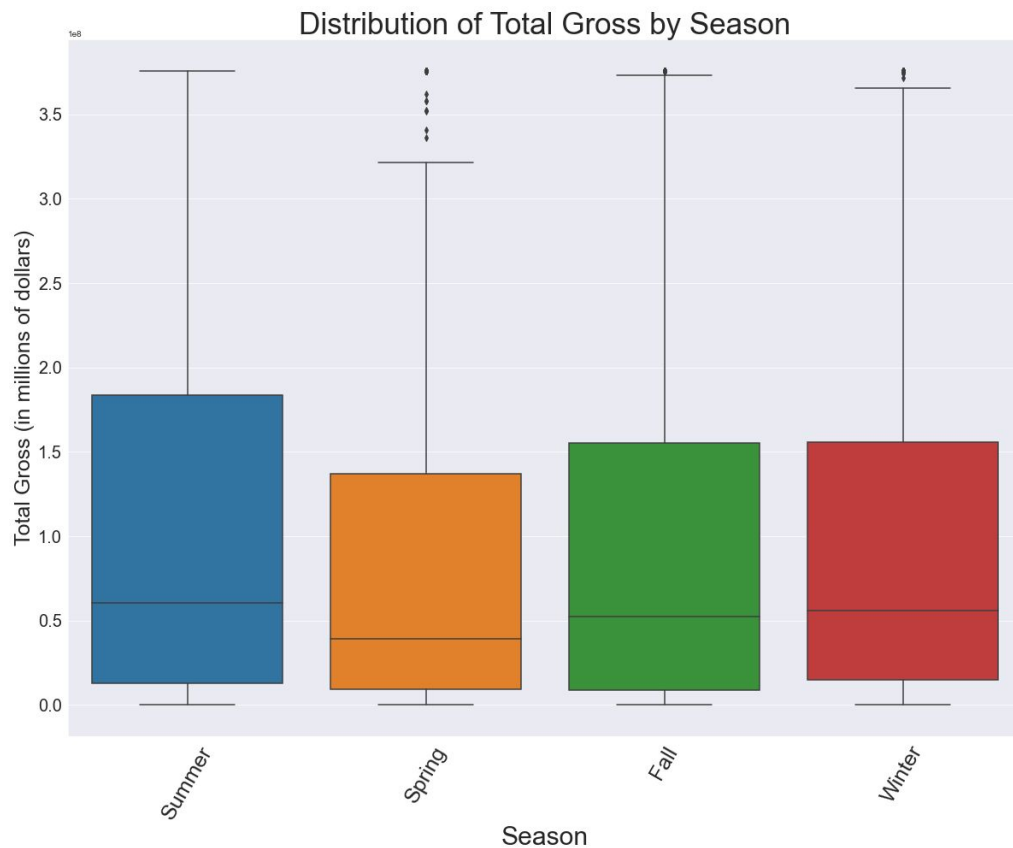
Distribution of Total Gross by Genre



(cut top 2%)

Still very large
range

1. Sci-Fi
2. Animation
3. Adventure
4. Action
5. Fantasy



(cut top 10%)

Looks relatively similar, but
remember outliers.

Summer Flic!

Conclusions

1. Doesn't have to be good
2. Lights, Camera, Action !
3. 🕶️

Further discussion:

-What are your parameters?

Thank You!

Email: hgcostello@gmail.com

GitHub: @hgcostello