Netflix Predictions

Mike Fiddler

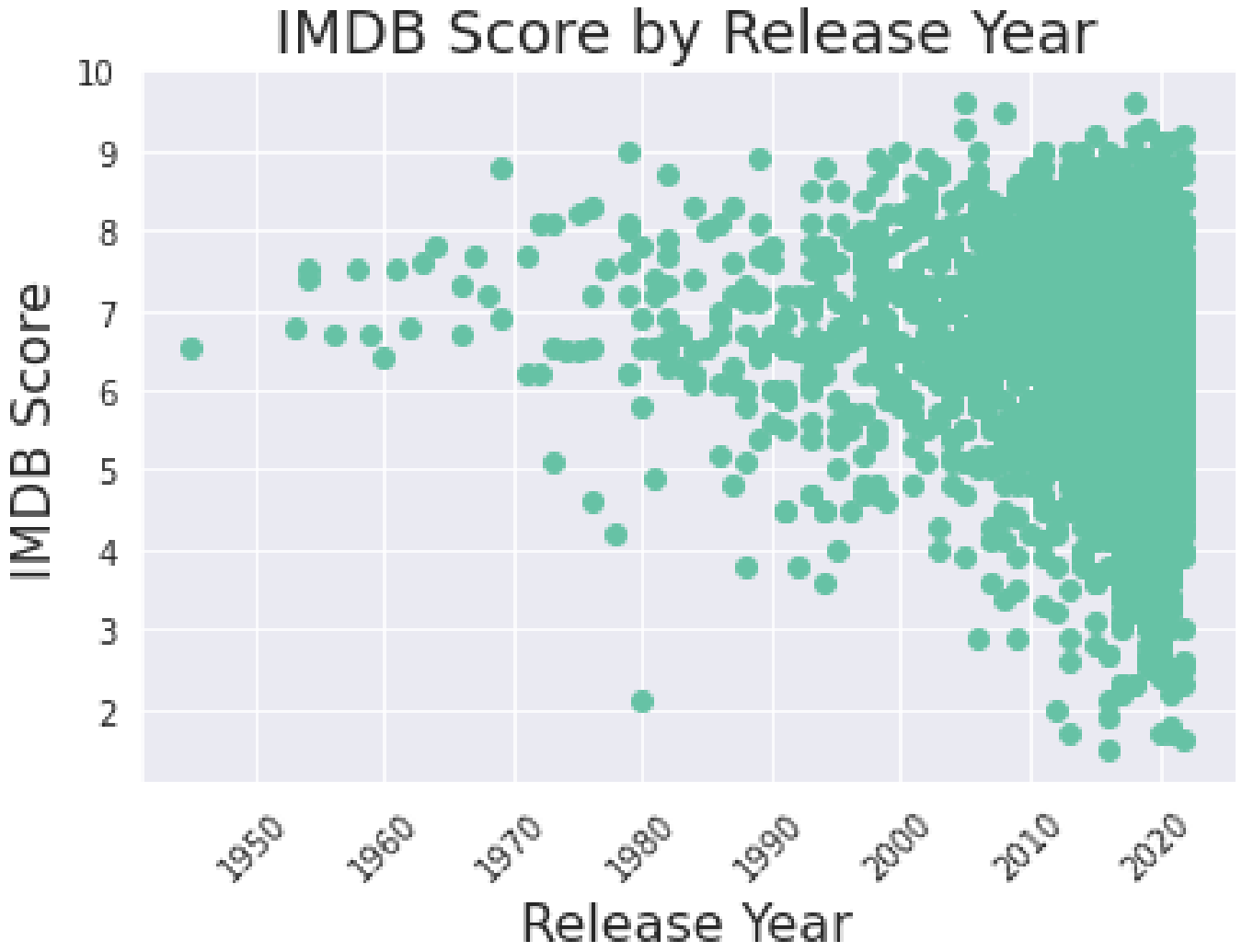
8/11/22

What should we produce to maximize profit?

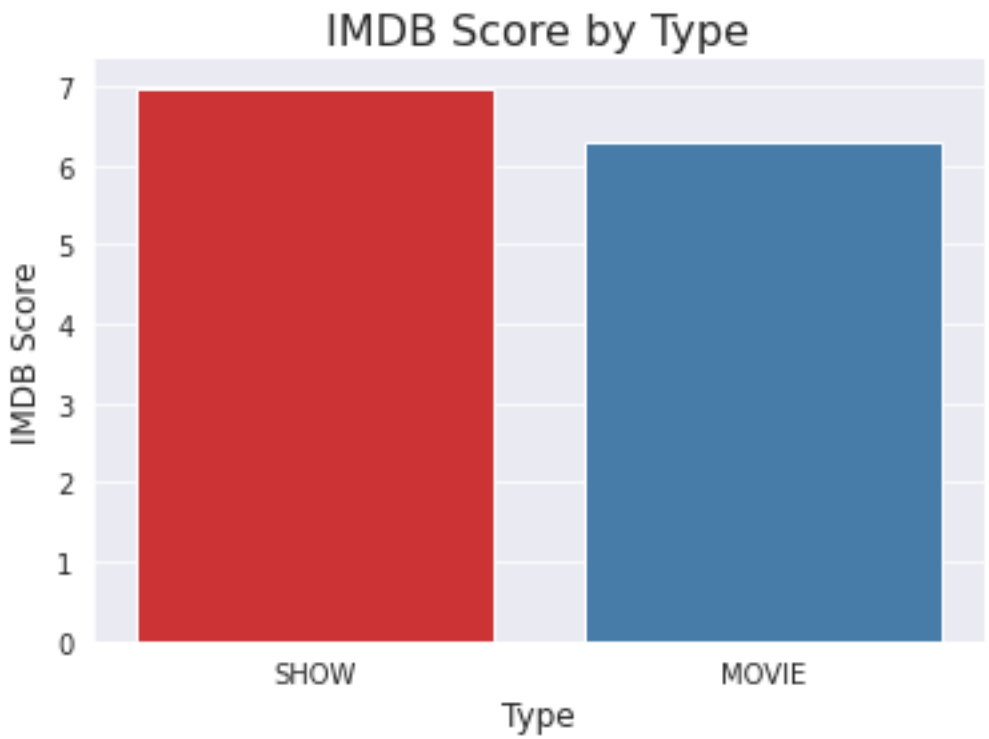
- We are looking at a data set from Netflix comprised of thousands of movies and shows to see if we can figure out what types of film we should be making for todays audiences.
- We are basing our predictions off of IMDB score. (A simple rating of 1-10, 10 being best and 1 being lowest score. IMDB is the Internet Movie Data Base.)
- Our features include type (show or movie), genre, TMDB score and runtime. (TMDB is The Movie Data Base a webpage similar to IMDB.

Relevance

- While our data goes back to 1940s most of our ratings and therefore predictions are of films over the past 20 years.



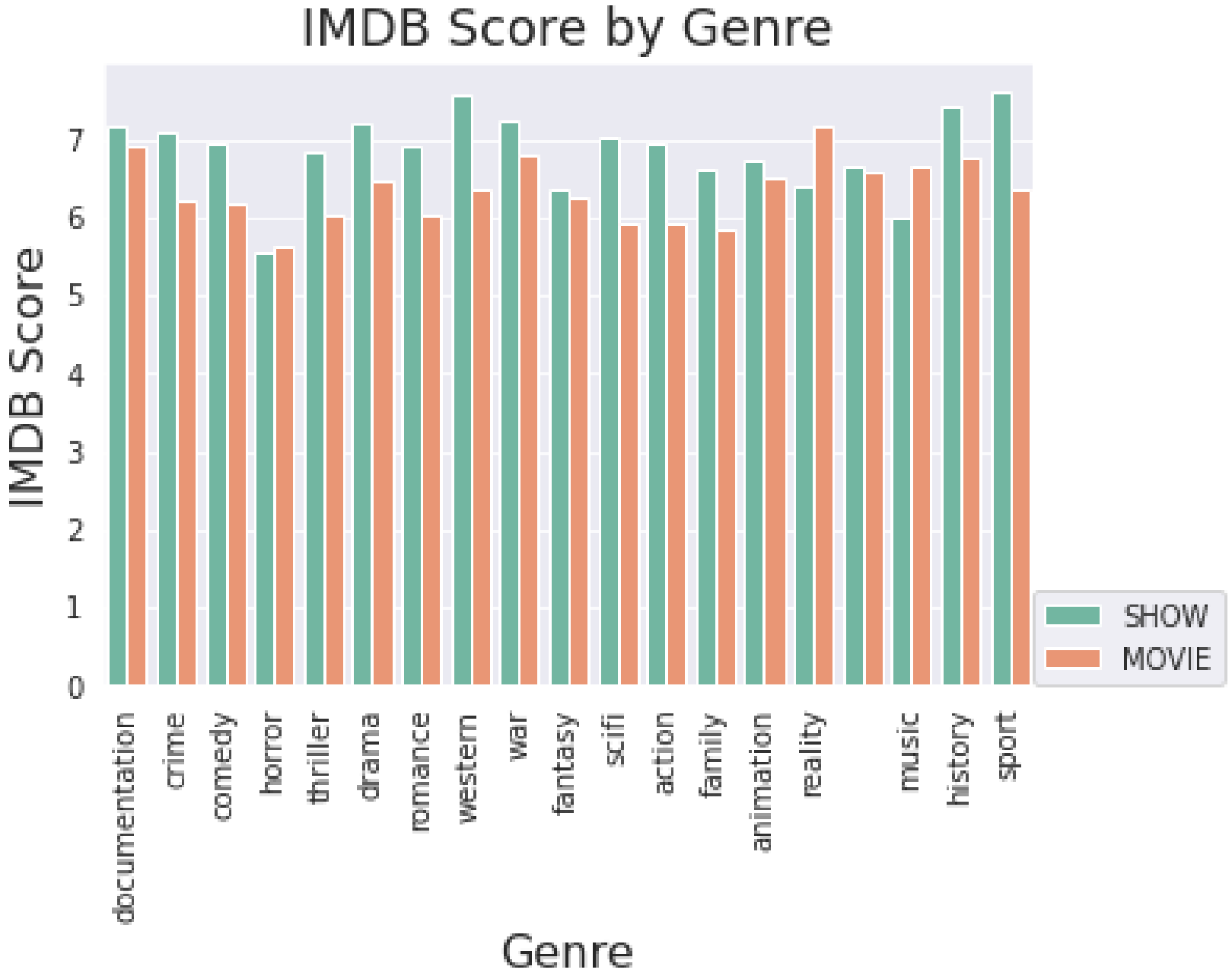
Is there a easy answer? No!



Here we see it’s not as easy as focusing on just one type of entertainment. While shows do get the edge in IMDB score we can see that people still love movies as well.

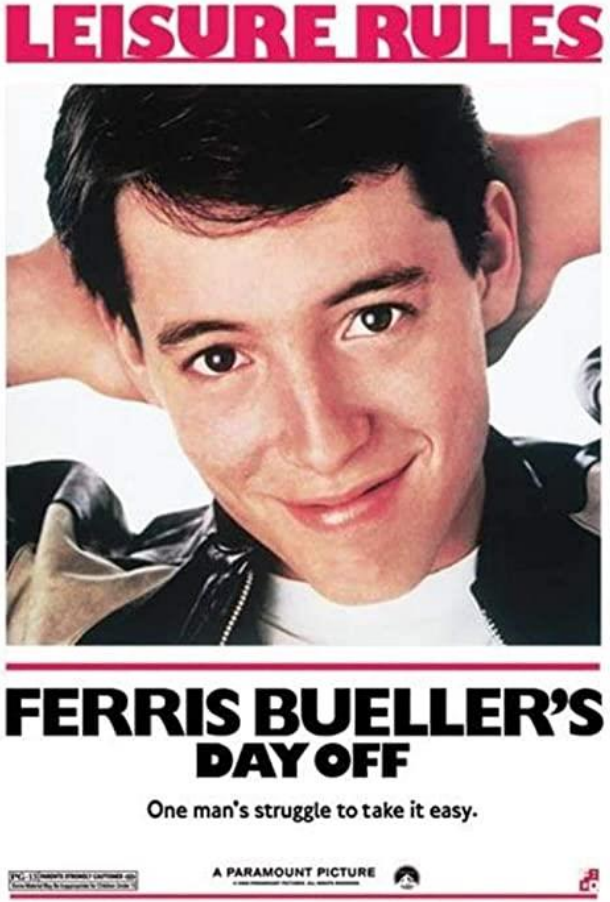
What to avoid!

- Horror genre over both types is the lowest.
- Scifi, action, western and family movies are also low scoring.



Limitations

- Films that become cult classics or are highly touted by critics does not translate into high IMDB scores. (Only IMDB uses can cast a vote on IMDB. This could include critics but their vote is not lent extra weight just because it's their job.)



Source: [Amazon.com: Ferris Buellers Day Off Leisure Rules One Mans Struggle to Take It Easy Comedy Movie Cool Wall Decor Art Print Poster Vintage Metal Tin Sign 12 x 8 Inch: Posters & Prints]

4 models down to 1

After pouring over all the metrics and ways I have to evaluate my models I recommend using the bagging model. While all my models are slightly overfit (they perform better on the training data vs the test data) I got my best results with the bagging model.

With this model we can look at a movie script or a pitch for a new tv show and know with confidence before we even spend a dime what we can expect that piece of film to rate with real world customers. With that knowledge we know what people like and what they will spend money on to see.

That’s the point, giving the people what they want to maximize profits on all future movies and shows.

Why am I confidant?

- My metrics prove that my model is accurate to within less then two thirds of a point on a base 10 scale.
- If a shows rating is 6 my machine is predicting between 5.4 and 6.5
- Even when my model does make a bigger error its prediction is still only off by .81
- See readme for more details about my metrics.

Thank You