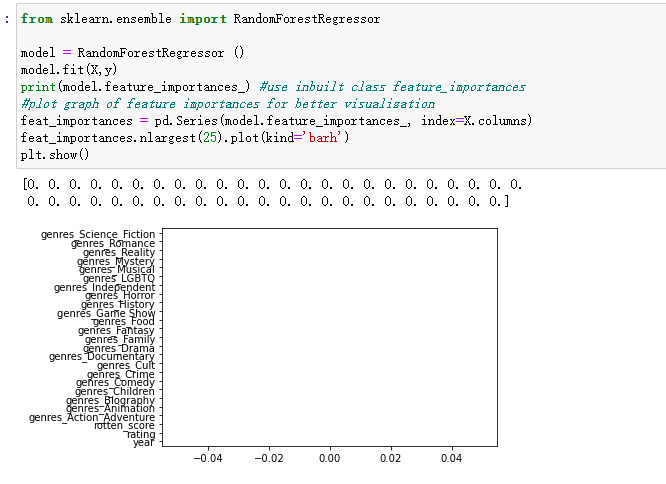
For imdb\_score, the mean is 6.794. We make it to categorical data, it is high when it > 6.794 and it is low when it < 6.794. We will predict whether it has high imdb score or low imdb score through machine learning models. A program that has high imdb score means high quality, and high quality programs are good for HBO.

Title and type are useless data and year is better than decade for ML model, so we delete them.

We do feature selection through built-in feature importance. No feature has high importance with imdb\_score. I choose all feature except rotten\_score because the function is similar with our target valuable imdb\_score.



The subset of our data : we use 80% data as the train subset



Feature columns(X1) are year, rating, genres\_action\_adventure, genres\_animation, genres\_biography, genres\_children, genres\_comedy, genres\_crime, genres\_cult, genres\_documentary, genres\_drama, genres\_family, genres\_fantasy, genres\_food, genres\_game\_show, genres\_history, genres\_horror, genres\_independent, genres\_lgbtq, genres\_musical, genres\_mystery, genres\_reality, genres\_romance, genres\_science\_fiction, genres\_sport, genres\_stand\_up\_talk, genres\_thriller, genres\_travel

Label column(y1) is imdb\_score