**Box Office Predictions**

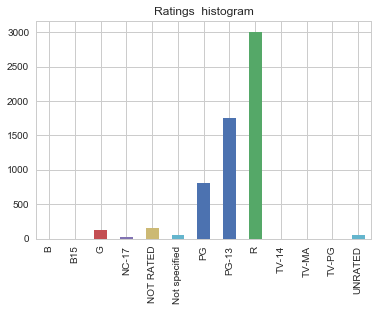
We’ll use interesting data such as a film’s budget, studio, director, and even its leading star to predict box office numbers before a movie hits the big screen. We’ll use movie data from 2014 and earlier to predict movies released in 2015 and 2016.

Data Shape:

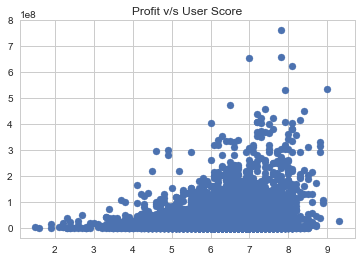
#6000 rows and 12 features

#no missing data

Class imbalance check:



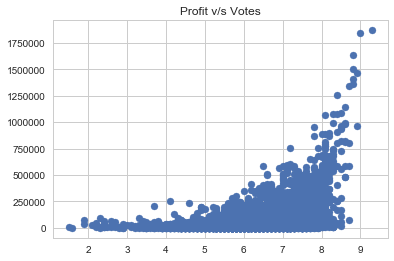
Check correlation between gross and score



Corr=0.16

Therefore, not correlated.

Check correlation between gross and votes



Corr:0.66

Somewhat correlated.

Check profit distribution

#mean gross=33416400; min=441; max=760508000; std=57352100

It is quite unbalanced.