

OLS

Linear Regression for "gross"

```
Call:
lm(formula = gross ~ ., data = i)

Residuals:
    Min       1Q   Median       3Q      Max
-321158186 -18624801 -5629817  9024931 467156928

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.754e+08  3.004e+08   0.917  0.35935
num_critic_for_reviews  8.106e+04  9.896e+03   8.191 3.43e-16 ***
duration      1.092e+05  3.917e+04   2.787  0.00534 **
director_facebook_likes -2.179e+03  2.880e+02  -7.568 4.65e-14 ***
actor_3_facebook_likes -1.105e+04  1.110e+03  -9.958 < 2e-16 ***
actor_1_facebook_likes -9.820e+03  6.663e+02 -14.739 < 2e-16 ***
num_voted_users  2.477e+02  1.068e+01  23.200 < 2e-16 ***
cast_total_facebook_likes  9.782e+03  6.641e+02  14.730 < 2e-16 ***
facenumber_in_poster  -7.332e+05  3.442e+05  -2.130  0.03323 *
num_user_for_reviews  1.860e+04  3.290e+03   5.653 1.69e-08 ***
budget       1.346e-02  3.269e-03   4.117 3.92e-05 ***
title_year   -1.254e+05  1.494e+05  -0.840  0.40121
actor_2_facebook_likes -9.175e+03  7.073e+02 -12.973 < 2e-16 ***
imdb_score   -6.133e+06  7.548e+05  -8.125 5.88e-16 ***
aspect_ratio  9.428e+04  7.483e+05   0.126  0.89974
movie_facebook_likes -8.876e+01  4.985e+01  -1.780  0.07507 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 44820000 on 4060 degrees of freedom
Multiple R-squared:  0.5514,    Adjusted R-squared:  0.5498
F-statistic: 332.7 on 15 and 4060 DF,  p-value: < 2.2e-16
```

Linear Regression result for imdb_score

```
Call:
lm(formula = imdb_score ~ ., data = training)

Residuals:
    Min       1Q   Median       3Q      Max
-4.4101 -0.4545  0.0640  0.5849  2.4289

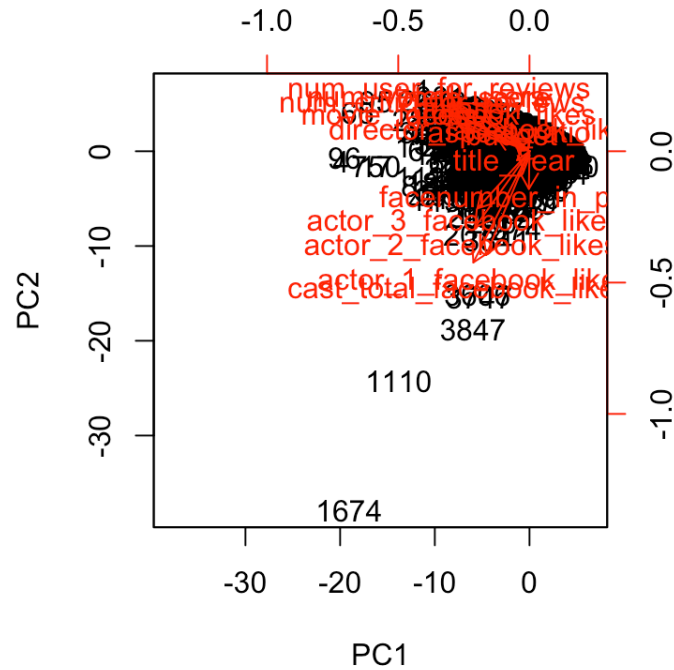
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.164e+01  8.196e+00   7.521 7.47e-14 ***
num_critic_for_reviews  2.582e-03  2.495e-04  10.349 < 2e-16 ***
duration      1.288e-02  9.453e-04  13.626 < 2e-16 ***
director_facebook_likes -3.131e-06  6.337e-06  -0.494 0.621333
actor_3_facebook_likes  8.259e-05  2.392e-05   3.453 0.000563 ***
actor_1_facebook_likes  7.451e-05  1.455e-05   5.119 3.29e-07 ***
gross       -2.421e-09  3.320e-10  -7.292 4.06e-13 ***
num_voted_users  4.030e-06  2.342e-07  17.204 < 2e-16 ***
cast_total_facebook_likes -7.382e-05  1.453e-05  -5.080 4.04e-07 ***
facenumber_in_poster  -2.999e-02  7.919e-03  -3.787 0.000156 ***
num_user_for_reviews  -7.275e-04  7.100e-05 -10.246 < 2e-16 ***
budget       -5.236e-11  6.445e-11  -0.812 0.416607
title_year   -2.838e-02  4.097e-03  -6.929 5.35e-12 ***
actor_2_facebook_likes  7.436e-05  1.533e-05   4.851 1.30e-06 ***
aspect_ratio  -3.947e-02  4.602e-02  -0.858 0.391155
movie_facebook_likes -1.542e-06  1.133e-06  -1.361 0.173650
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8542 on 2555 degrees of freedom
(723 observations deleted due to missingness)
Multiple R-squared:  0.3422,    Adjusted R-squared:  0.3383
F-statistic: 88.6 on 15 and 2555 DF,  p-value: < 2.2e-16
```

PCA

with 2 Principle component (should be more, but muti-D plots are hard to display)

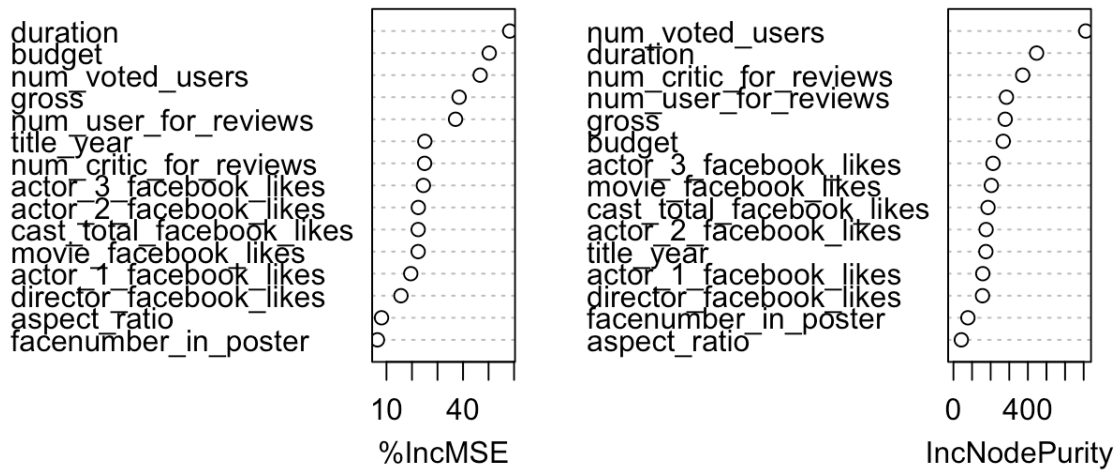
We can see face_book_like variables could be combined in to one variable.



Random Forest

Importance score for imdb_scores

score.rf



Importance scores for gross

gross.rf

