

Feature Target Correlation Analysis

Positive Correlation with US Gross

| | |
|-----------|------|
| Duration | 0.56 |
| Budget | 0.72 |
| Action | 0.42 |
| Adventure | 0.49 |
| Sci-Fi | 0.24 |
| PG | 0.42 |
| PG-13 | 0.33 |

Negative Correlation with US Gross

| | |
|-----------|--------|
| R Rating | - 0.47 |
| Biography | -0.11 |
| Comedy | -0.03 |
| Crime | -0.25 |
| Drama | -0.21 |
| Fantasy | -0.08 |
| History | -0.03 |
| Horror | -0.36 |
| Musical | -0.06 |
| Mystery | -0.21 |
| Romance | -0.14 |
| Sport | -0.09 |
| War | -0.10 |
| Western | -.07 |

Close to zero

| | |
|-----------|--------|
| Family | 0.01 |
| Music | 0.18 |
| Animation | - 0.00 |

All included in the linear regression model

Coefficients

Duration 1,647,406.25

Budget 0.39

Action -20,658,701.63

Adventure 19,793,585.00

Sci-Fi 19,450,110.29

PG 68,017,516.47

PG-13 -27,713,940.97

R Rating -40,303,575.50

Biography -57,640,257.49

Comedy 12,575,077.35

Crime -10,495,540.63

Drama -28,094,114.77

Fantasy -7,649,883.02

History -76,092,866.62

Horror -14,634,199.66

Musical -67,808,047.35

Mystery -12,610,294.09

Romance -29,425,524.70

Sport -0.00

War 0.00

Western 0.00

Family 0.00

Music -0.0

Animation -54,468,082.93

R^2

Slight overfitting because Train is more than test

Train R^2 .318

Test R^2 .276

MAE: Mean Absolute Error

How close the prediction is against the real value

\$46,825,271