IGN Video Game Reviews

Machine Learning

by Ellen A Savoye

July 9, 2018

Predictive Models

Subsetting the Data

Prior to any ventures into modeling, a data set was created to remove any unnecessary features and any blank records. 36 records with a 'blank' genre were removed. Given that some of the kept features are categorical, a conversion to binary/dummy variables was needed using model matrix. I chose to not combine release day, month, and year into a combined date field as I don't believe it would have given any valuable insight as unique a combined value.

```
##   score_phrase        title      platform         score          genre
##            0            0            0            0            36
## editors_choice   release_year  release_month    release_day  platform_group
##            0            0            0            0            0
##    genre_group
##           36

## 'data.frame':    18589 obs. of  7 variables:
##  $ score        : num  9 9 8.5 8.5 8.5 7 3 9 3 7 ...
##  $ editors_choice: Factor w/ 2 levels "N","Y": 2 2 1 1 1 1 1 2 1 1 ...
##  $ release_year : Factor w/ 21 levels "1996","1997",..: 17 17 17 17 17 17 17 17 17 17 ...
##  $ release_month : Factor w/ 12 levels "1","2","3","4",..: 9 9 9 9 9 9 9 9 9 9 ...
##  $ release_day  : Factor w/ 31 levels "1","2","3","4",..: 12 12 12 11 11 11 11 11 11 11 ...
##  $ platform_group: Factor w/ 14 levels "Android","Apple",..: 9 9 2 14 9 2 14 12 9 12 ...
##  $ genre_group  : Factor w/ 21 levels "Action","Adventure",..: 12 12 14 19 19 20 6 16 6 20 ...

## [1] 18589     76

##  [1]  1  4  6  7 11 12 14 17 18 19 20 21 22 23 24 25 26 27 28 32 35 36 37
## [24] 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
## [47] 61 62 63 64 65

## [1] 18589     25

## integer(0)

## reg_data
##
##  26  Variables      18589  Observations
## ---------------------------------------------------------------------------
## editors_choiceY
##       n  missing distinct     Info     Sum     Mean      Gmd
##    18589        0        2     0.46     3515   0.1891   0.3067
##
## ---------------------------------------------------------------------------
## platform_groupApple
```

```
##        n  missing distinct     Info     Sum     Mean      Gmd
##    18589        0        2    0.156    1025  0.05514   0.1042
##
## -------------------------------------------------------------------------------
## platform_groupGame Boy
##        n  missing distinct     Info     Sum     Mean      Gmd
##    18589        0        2    0.153    1001  0.05385   0.1019
##
## -------------------------------------------------------------------------------
## platform_groupNintendo
##        n  missing distinct     Info     Sum     Mean      Gmd
##    18589        0        2    0.497    3896   0.2096   0.3313
##
## -------------------------------------------------------------------------------
## platform_groupOther
##        n  missing distinct     Info     Sum     Mean      Gmd
##    18589        0        2    0.147     959  0.05159  0.09786
##
## -------------------------------------------------------------------------------
## platform_groupPlayStation
##        n  missing distinct     Info     Sum     Mean      Gmd
##    18589        0        2    0.594    5055   0.2719    0.396
##
## -------------------------------------------------------------------------------
## platform_groupWindows
##        n  missing distinct     Info     Sum     Mean      Gmd
##    18589        0        2    0.447    3383    0.182   0.2978
##
## -------------------------------------------------------------------------------
## platform_groupXbox
##        n  missing distinct     Info     Sum     Mean      Gmd
##    18589        0        2    0.368    2660   0.1431   0.2453
##
## -------------------------------------------------------------------------------
## genre_groupAdventure
##        n  missing distinct     Info     Sum     Mean      Gmd
##    18589        0        2    0.191    1267  0.06816    0.127
##
## -------------------------------------------------------------------------------
## genre_groupRacing
##        n  missing distinct     Info     Sum     Mean      Gmd
##    18589        0        2    0.189    1258  0.06767   0.1262
##
## -------------------------------------------------------------------------------
## genre_groupRPG
##        n  missing distinct     Info     Sum     Mean      Gmd
##    18589        0        2    0.165    1089  0.05858   0.1103
##
## -------------------------------------------------------------------------------
## genre_groupShooter
##        n  missing distinct     Info     Sum     Mean      Gmd
##    18589        0        2    0.238    1614  0.08683   0.1586
##
## -------------------------------------------------------------------------------
```

```
## genre_groupSports
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2    0.287     1988   0.1069    0.191
##
## --------------------------------------------------------------------------
## genre_groupStrategy
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2    0.163     1071  0.05761   0.1086
##
## --------------------------------------------------------------------------
## release_month2
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2    0.199     1327  0.07139   0.1326
##
## --------------------------------------------------------------------------
## release_month3
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2    0.231     1565  0.08419   0.1542
##
## --------------------------------------------------------------------------
## release_month4
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2     0.19     1263  0.06794   0.1267
##
## --------------------------------------------------------------------------
## release_month5
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2    0.173     1141  0.06138   0.1152
##
## --------------------------------------------------------------------------
## release_month6
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2     0.22     1481  0.07967   0.1467
##
## --------------------------------------------------------------------------
## release_month7
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2    0.179     1188  0.06391   0.1197
##
## --------------------------------------------------------------------------
## release_month8
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2      0.2     1334  0.07176   0.1332
##
## --------------------------------------------------------------------------
## release_month9
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2    0.249     1701  0.09151   0.1663
##
## --------------------------------------------------------------------------
## release_month10
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2    0.326     2306   0.1241   0.2173
##
```

```
## ------------------------------------------------------------------------
## release_month11
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2    0.367     2655   0.1428   0.2449
##
## ------------------------------------------------------------------------
## release_month12
##        n  missing distinct     Info      Sum     Mean      Gmd
##    18589        0        2    0.223     1504  0.08091   0.1487
##
## ------------------------------------------------------------------------
## score
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##    18589        0       93    0.998    6.951    1.889      3.5      4.5
##      .25      .50      .75      .90      .95
##      6.0      7.3      8.2      8.9      9.1
##
## lowest :  0.5  0.7  0.8  1.0  1.1, highest:  9.6  9.7  9.8  9.9 10.0
## ------------------------------------------------------------------------
```

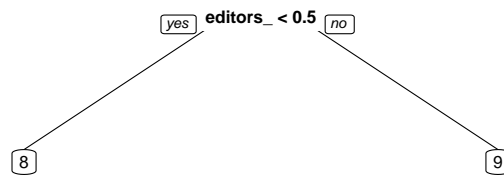**Linear Regression**

```
##
## Call:
## lm(formula = score ~ ., data = reg_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9827 -0.6511  0.2158  1.0260  3.4121
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                6.61710    0.07086  93.378  < 2e-16 ***
## editors_choiceY            2.29780    0.02693  85.334  < 2e-16 ***
## platform_groupApple        0.04548    0.07301   0.623 0.533318
## platform_groupGame.Boy    -0.63079    0.07332  -8.603  < 2e-16 ***
## platform_groupNintendo    -0.58438    0.06221  -9.393  < 2e-16 ***
## platform_groupOther       -0.33464    0.07393  -4.526 6.05e-06 ***
## platform_groupPlayStation -0.33149    0.06135  -5.403 6.63e-08 ***
## platform_groupWindows     -0.20208    0.06353  -3.181 0.001470 **
## platform_groupXbox        -0.20217    0.06425  -3.147 0.001654 **
## genre_groupAdventure       0.05884    0.04268   1.379 0.168006
## genre_groupRacing         -0.15182    0.04251  -3.572 0.000356 ***
## genre_groupRPG             0.42068    0.04576   9.194  < 2e-16 ***
## genre_groupShooter         0.07990    0.03853   2.074 0.038113 *
## genre_groupSports          0.04943    0.03503   1.411 0.158236
## genre_groupStrategy        0.21563    0.04755   4.535 5.80e-06 ***
## release_month2             0.16384    0.05764   2.843 0.004478 **
## release_month3             0.16889    0.05565   3.035 0.002412 **
## release_month4             0.15263    0.05829   2.619 0.008837 **
## release_month5             0.11642    0.05978   1.947 0.051499 .
## release_month6             0.22268    0.05622   3.961 7.51e-05 ***
## release_month7             0.12258    0.05923   2.069 0.038518 *
## release_month8             0.26927    0.05757   4.678 2.92e-06 ***
```

```
## release_month9            0.35055     0.05473   6.406 1.53e-10 ***
## release_month10           0.26770     0.05179   5.169 2.38e-07 ***
## release_month11           0.19154     0.05070   3.777 0.000159 ***
## release_month12           0.02856     0.05603   0.510 0.610195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.42 on 18563 degrees of freedom
## Multiple R-squared:  0.3132, Adjusted R-squared:  0.3123
## F-statistic: 338.6 on 25 and 18563 DF,  p-value: < 2.2e-16
```

Previous iterations of the linear regression showed multicollinearity between score and score phrase with a multiple R-squared of 0.9726. After removing score phrase and re-running, R-squared drops to 0.3187.

Logistic Regression

Random Forest


```
##
## Call:
##  randomForest(formula = score ~ ., data = Train, ntree = 500)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 8
##
##           Mean of squared residuals: 2.041761
##                     % Var explained: 30.43
```

```
## [1] 0.3198914

##           [,1]
## 0.5 0.3039893
## 0.7 0.3039893
## 0.8 0.3039893
## 1   0.3039893
## 1.1 0.3039893
## 1.2 0.3039893
## 1.3 0.3039893
## 1.4 0.3039893
## 1.5 0.3039893
## 1.7 0.3039893
## 1.8 0.3039893
## 1.9 0.3039893
## 2   0.3039893
## 2.1 0.3039893
## 2.2 0.3039893
## 2.3 0.3039893
## 2.4 0.3039893
## 2.5 0.3039893
## 2.6 0.3039893
## 2.7 0.3039893
## 2.8 0.3039893
## 2.9 0.3039893
## 3   0.3039893
## 3.1 0.3039893
## 3.2 0.3039893
## 3.3 0.3039893
## 3.4 0.3039893
## 3.5 0.3039893
## 3.6 0.3039893
## 3.7 0.3039893
## 3.8 0.3039893
## 3.9 0.3039893
## 4   0.3039893
## 4.1 0.3039893
## 4.2 0.3039893
## 4.3 0.3039893
## 4.4 0.3039893
## 4.5 0.3039893
## 4.6 0.3039893
## 4.7 0.3039893
## 4.8 0.3039893
## 4.9 0.3039893
## 5   0.3039893
## 5.1 0.3039893
## 5.2 0.3039893
## 5.3 0.3039893
## 5.4 0.3039893
## 5.5 0.3039893
## 5.6 0.3039893
## 5.7 0.3039893
## 5.8 0.3039893
```

```
## 5.9 0.3039893
## 6   0.3039893
## 6.1 0.3039893
## 6.2 0.3039893
## 6.3 0.3039893
## 6.4 0.3039893
## 6.5 0.3039893
## 6.6 0.3039893
## 6.7 0.3039893
## 6.8 0.3039893
## 6.9 0.3039893
## 7   0.3039893
## 7.1 0.3039893
## 7.2 0.3039893
## 7.3 0.3039893
## 7.4 0.3039893
## 7.5 0.3039893
## 7.6 0.3039893
## 7.7 0.3039893
## 7.8 0.3039893
## 7.9 0.3039893
## 8   0.3039893
## 8.1 0.3039893
## 8.2 0.3039893
## 8.3 0.3039893
## 8.4 0.3039893
## 8.5 0.3039893
## 8.6 0.3039893
## 8.7 0.3039893
## 8.8 0.3039893
## 8.9 0.3039893
## 9   0.3039893
## 9.1 0.3039893
## 9.2 0.3039893
## 9.3 0.3039893
## 9.4 0.3039893
## 9.5 0.3039893
## 9.6 0.3039893
## 9.7 0.3039893
## 9.8 0.3039893
## 9.9 0.3039893
## 10  0.3039893
```

Conclusion

After looking at platform, genre, and release date in a myriad of ways, Action can be conclusively named the top pick in Genre for types of video games released. When looking at Genre by score, RPG is the favorite when it comes to review score. When looking at the data by Platform, the answer is not as clear. While Playstation dominated for all years combined, even though it was released years after Nintendo or Windows, the exhibit by score does not show Playstation to have the highest rated games. This would imply that while a platform can release a large amount of games, it does not necessarily mean that those games will be favorably reviewed. To truly determine what platform reigns supreme, if there is one, I would look into tying

video game sale data with the video game reviews to look at sale volume versus platform, genre, date, and score.