Game Studio Analytics

Team 9

Ravikiran Pise, Tyler Scott, Madeeha Haryani, Tong Shi, Rishabh Sachdeva

What business problem are you addressing?

Strategizing

Identify the type of titles that succeeded and what was the relation between the markets and users

Title Success

How did the titles perform on different consoles and what was the correlation between user and critic ratings

Choice of Platform

To grasp the relationship between the success of a title and the platform it was launched on so that a recommendation to initiate the new product development for the right platform

Porter's Five Forces for the industry

Bargaining Power of Suppliers

High

Suppliers involved are Game Engine developing firms, infrastructure firms, etc and because of high customer volumes, the bargaining power is considered high

Bargaining Power of Customers

High

Customers have an array of options to choose from and hence this factor is considered high

Threat of New Entrants

High

There have been more than 10000 game studios that have launched live games and hence threat remains High

Availability of Substitutes

High

Substitute products are available at ease and this makes the factor to be marked at High

Intensity of Competition

High

Gaming industry has evolved to be a place of high competition and direct rivals like EA, Ubisoft, SONY, and Nintendo have made the classification to be High

Wargaming Studios

Key Facts

- 3 key titles – World of Tanks ($400 M), World of Warships (< $20 M), World of Warplanes
- ~2000 employees located in Europe (Russia, UK, Lithuania, Cyprus), US, & China
- Global Player base – WoT (160 M), WoWS (20 M), WoWP (5 M). Major concentrations in EU and RUS servers
- WoT and WoWS have been active since 2010 and 2013 respectively
- Studio has shown dedication to revamp the titles as the industry's landscape has changed over time

Shrinking Player base

Games have been losing player base to rival games.

N(players leaving) > N(new players)

Pegged by a Single game

Major chunk of revenue is dependent on WoT (> 90%)

The audience has been moving to World-building and Action RPG games and a title in these genres would help diversify the revenue streams of the studio

Affected by RUS-UKR conflict

Most of the ACTIVE player base of the game is in RUS and EU regions.

Previously headquartered in Belarus, the studio had to sell the operations of their RUS region to a 3rd party and take a major revenue hit

SWOT Analysis of WG

STRENGTHS

A Loyal and large customer-base
Historically successful
Adaptable to trends

S

WEAKNESSES

W

Dependent on a single game
Same kind of products
Affected by external factors

O

Emergence of new platforms
High potential in Asian markets

OPPORTUNITIES

T

Highly competitive market
Shrinking user base
High availability of substitutes

THREATS

Dataset - Video Game Sales and Ratings
Source - Kaggle

Link to the Dataset

Data Description

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Name	Platfor	ar_of_Re	Genre	Publish	NA_Sal	EU_Sal	JP_Sale	Other_S	Global_S	Critic_Sco	User_Sco	Avg_Sco	Succes
2	The Elder Scrolls	PC	2014	Role-Playin	Bethesda Sc	0.33	0.54	0	0.07	0.94	7.1	5.7	6.4	0
3	Call of Duty: Wo	Wii	2008	Shooter	Activision	1.17	0.58	0	0.18	1.94	8.3	7.6	8.0	1
4	Crayon Shin-Cha	GBA	2006	Action	Banpresto	0	0	0.07	0	0.07	4.9	8.2	6.6	0
5	Harvey Birdman:	PS2	2008	Adventure	Capcom	0.06	0.04	0	0.01	0.12	6.3	4.4	5.4	0
6	Iza, Shutsujin! Ko	PSP	2011	Adventure	Quinrose	0	0	0.01	0	0.01	5.2	4.2	4.7	0
7	Cartoon Networ	DS	2006	Action	Game Facto	0.04	0	0	0	0.05	5.5	6.3	5.9	0
8	Ninokuni: Shikko	DS	2010	Role-Playin	Level 5	0	0	0.54	0	0.54	6.2	7.1	6.7	1

Dependent:
Success

Independent:
Global_Sales
Platform
Genre
Avg_Score

Data is considered for games post 2005

Representation:
Each row represents the Name, Platform, Year of Release, Genre, Publisher, Sales in major Markets (NA, EU, JP), Global Sales, Critic Score, User Score, and Success of a game.

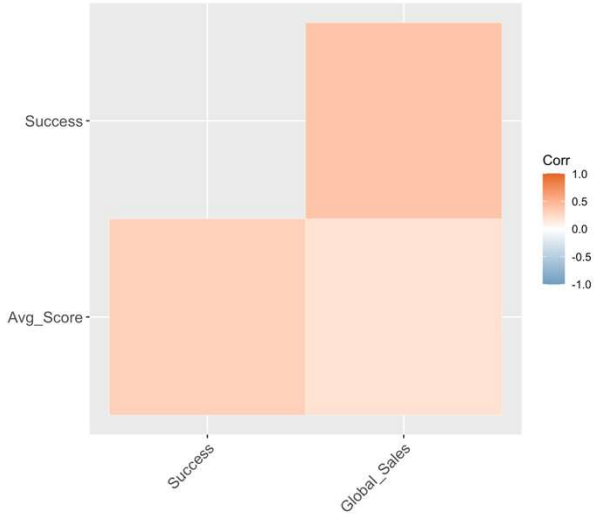
Sales columns are a measure of copies sold and represented in Millions

Critic Score and User Score are parameters gauged out of 10 and an Avg. score is the mean of the values

Success is a binary parameter with 0 representing failure and 1 representing success

Correlation

```
> print(ibcf)
              Success Global_Sales Avg_Score
Success     1.0000000    0.3756899 0.3079157
Global_Sales 0.3756899    1.0000000 0.1955340
Avg_Score   0.3079157    0.1955340 1.0000000
```



- In this dataset, the correlation between just the user and critics' scores and global sales is quite low
- This led us to believe that while the way games were being scored were still important factors when deciding on the next game release, it would not be sufficient
- For example, a game in our dataset like Metal Gear Solid 3: Subsistence, has been praised by both fans and critics for its quality, but did not sell as well as other titles that scored significantly less
- Scores alone are not the best deterrent for potential success

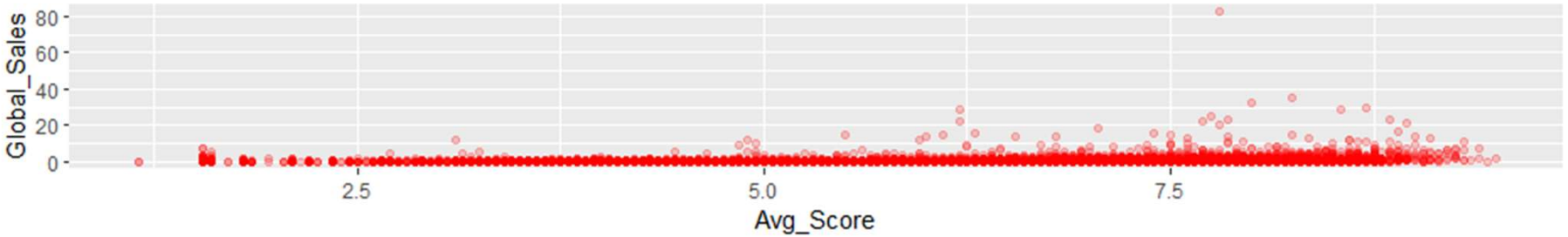
Data Summarization

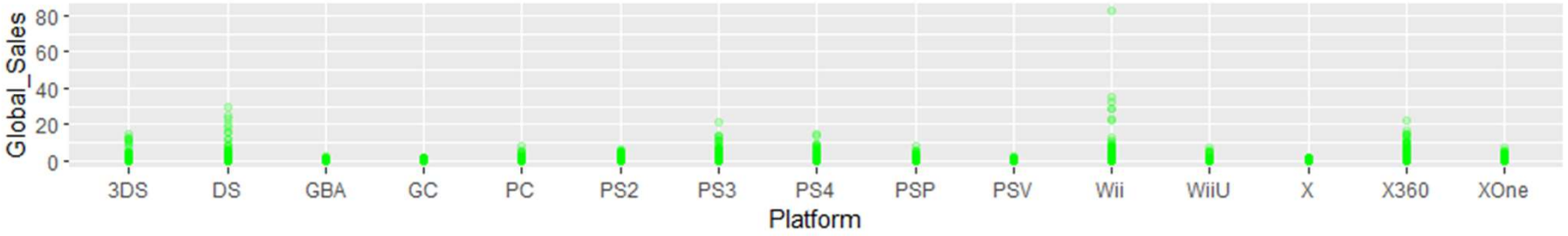
	mean	median	sd	variance	min	max	count	miss.val
Global Sales	0.4711569	0.13	1.533623	2.352000	0.01	82.54	11747	0
Avg Score	5.7182770	6.25	2.050531	4.204679	1.15	9.50	11747	0

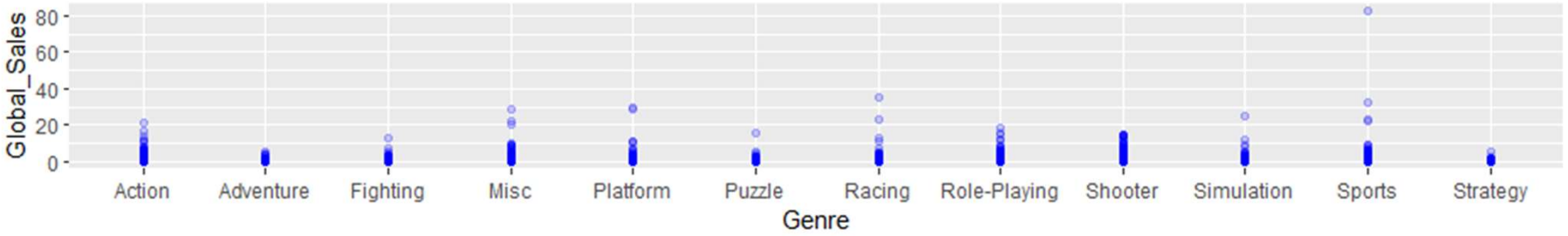
	3DS	DS	GBA	GC	PC	PS2	PS3	PS4	PSP	PSV	Wii	WiiU	X	X360	XOne
Action	196	338	42	37	175	190	383	152	218	168	237	64	57	324	88
Adventure	34	244	10	9	75	170	72	29	240	117	83	4	12	47	16
Fighting	15	36	0	11	5	82	78	20	71	16	43	5	11	67	7
Misc	56	384	41	12	27	142	123	25	103	25	281	22	14	124	20
Platform	29	90	31	20	11	38	36	15	37	8	58	18	14	25	5
Puzzle	20	236	10	1	29	7	3	2	43	3	56	4	0	7	0
Racing	11	65	10	8	57	72	91	19	67	11	99	4	29	106	20
Role-Playing	91	204	15	4	101	90	119	52	196	85	36	7	5	77	16
Shooter	7	42	3	9	140	57	155	41	39	5	65	10	51	203	38
Simulation	32	296	6	4	106	33	33	6	30	4	87	1	6	40	5
Sports	27	139	12	26	57	167	213	47	132	23	253	8	46	224	39
Strategy	16	81	2	2	155	36	24	7	63	9	25	3	5	28	4

Boxplots

Scatterplots







Data Modeling

- Decision Trees
- Random Forest
- Logistic Regression
- Multiple Linear Regression

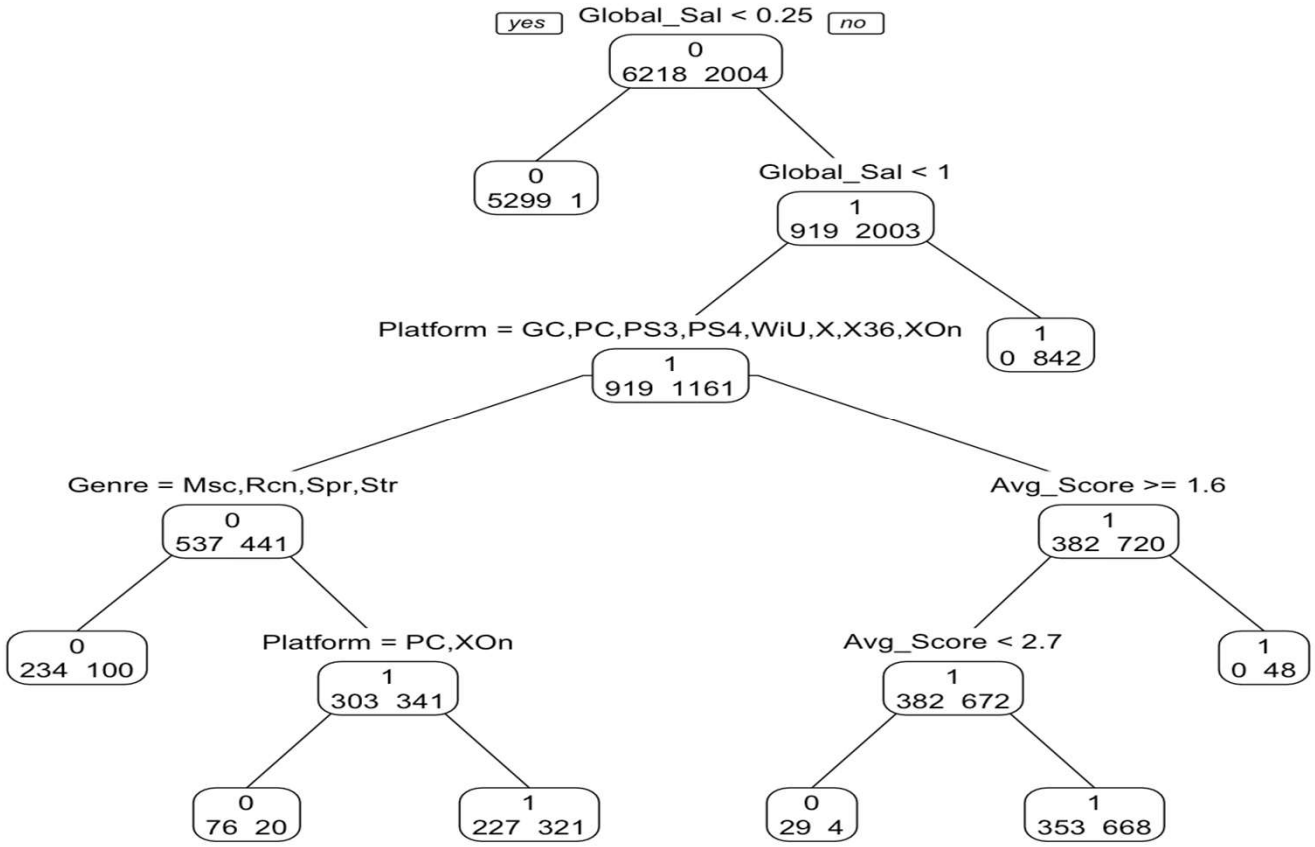
Decision Tree

- Target Variable – Success
- Prediction – To find predictive probability of the game being Successful or Not.
- Independent Variables – Global_Sales, Platform, Genre, and Avg_Score
- Predictive Accuracy – 91.74%

```
> decisiontree.prediction <- data.frame(Global_Sales = 0.6, Genre = 'Role-Playing', Platform='PC', Avg_Score=2)
> predict(videogame.decisiontree, decisiontree.prediction, method = "anova", type='prob')
          0         1
1 0.7916667 0.2083333
> decisiontree.prediction2 <- data.frame(Global_Sales = 0.2, Genre = 'Role-Playing', Platform='PC', Avg_Score=9)
> predict(videogame.decisiontree, decisiontree.prediction2, method = "anova", type='prob')
          0           1
1 0.9998113 0.0001886792
```

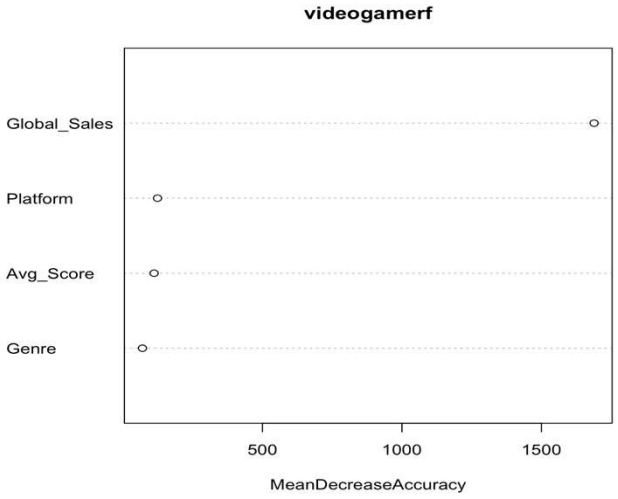
```
> confusionMatrix(decisiontree.pred, as.factor(valid.df$Success))
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2441   45
         1  246  793

               Accuracy : 0.9174
                 95% CI : (0.9079, 0.9263)
    No Information Rate : 0.7623
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.7896

 Mcnemar's Test P-Value : < 0.00000000000000022

            Sensitivity : 0.9084
            Specificity : 0.9463
         Pos Pred Value : 0.9819
         Neg Pred Value : 0.7632
             Prevalence : 0.7623
         Detection Rate : 0.6925
   Detection Prevalence : 0.7052
      Balanced Accuracy : 0.9274

       'Positive' Class : 0

> decisiontree.imp <- as.data.frame(videogame.decisiontree$variable.importance)
> decisiontree.imp
             videogame.decisiontree$variable.importance
Global_Sales                                  2006.02403
Avg_Score                                      398.61355
Platform                                       150.50831
Genre                                           58.30862
```

Decision Tree



Random Forest

- Target Variable – Success
- Prediction – To predict the most Important Variables.
- Independent Variables – Global_Sales, Platform, Genre, and Avg_Score
- Predictive Accuracy – 91.49%




```
> videogamerf <- randomForest(as.factor(Success) ~ Global_Sales + Genre + Platform + Avg_Score, data = train.df,
+                             ntree = 5000, nodesize = 1, importance = TRUE, sampsize = 8000)
> confusionMatrix(videogamerf.pred, valid.df$Success)
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2487  100
         1  200  738

               Accuracy : 0.9149
                 95% CI : (0.9052, 0.9239)
    No Information Rate : 0.7623
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.7744

 Mcnemar's Test P-Value : 0.00000001092

            Sensitivity : 0.9256
            Specificity : 0.8807
         Pos Pred Value : 0.9613
         Neg Pred Value : 0.7868
             Prevalence : 0.7623
         Detection Rate : 0.7055
   Detection Prevalence : 0.7339
      Balanced Accuracy : 0.9031

       'Positive' Class : 0
```

Logistic Regression

- Target Variable – Success
- Prediction – To predict the most Successful Platform and Genre.
- Independent Variables – Global_Sales, Platform, Genre, and Avg_Score
- Predictive Accuracy – 91.49%

```
> confusionMatrix(table(predict(videogamelogistic.reg, newdata = valid.df,
+                               type="response") >= 0.5, valid.df$Success == 1))
Confusion Matrix and Statistics


       FALSE TRUE
  FALSE  2554  310
  TRUE    133  528

               Accuracy : 0.8743
                 95% CI : (0.8629, 0.8851)
    No Information Rate : 0.7623
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.6261

 Mcnemar's Test P-Value : < 0.00000000000000022

            Sensitivity : 0.9505
            Specificity : 0.6301
         Pos Pred Value : 0.8918
         Neg Pred Value : 0.7988
             Prevalence : 0.7623
         Detection Rate : 0.7245
   Detection Prevalence : 0.8125
      Balanced Accuracy : 0.7903

       'Positive' Class : FALSE
```

```
Call:
glm(formula = Success ~ Global_Sales + Genre + Platform + Avg_Score,
    family = "binomial", data = train.df)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.69558  -0.40529  -0.26821  -0.06872   2.99938

Coefficients:
                  Estimate Std. Error z value             Pr(>|z|)
(Intercept)       -3.453280   0.230944 -14.953 < 0.0000000000000002 ***
Global_Sales       6.009126   0.172698  34.796 < 0.0000000000000002 ***
GenreAdventure    -0.817615   0.187773  -4.354   0.0000133515416466 ***
GenreFighting     -0.149593   0.191816  -0.780            0.435465
GenreMisc         -0.561586   0.138911  -4.043   0.0005282102343600 ***
GenrePlatform     -0.232616   0.195554  -1.190            0.234234
GenrePuzzle       -1.151157   0.290858  -3.958   0.0000756447983129 ***
GenreRacing       -0.709490   0.190398  -3.726            0.000194 ***
GenreRole-Playing -0.379358   0.149055  -2.545            0.010925 *
GenreShooter      -0.115557   0.160187  -0.721            0.470670
GenreSimulation   -0.085924   0.174419  -0.493            0.622274
GenreSports       -1.119524   0.146181  -7.659   0.0000000000000188 ***
GenreStrategy     -0.823724   0.268379  -3.069            0.002146 **
PlatformDS         0.028271   0.204473   0.138            0.890033
PlatformGBA        0.263778   0.339268   0.777            0.436870
PlatformGC        -0.147460   0.381322  -0.387            0.698974
PlatformPC        -2.265812   0.333837  -6.787   0.0000000000114346 ***
PlatformPS2        0.452242   0.213536   2.118            0.034186 *
PlatformPS3       -0.428859   0.215937  -1.986            0.047029 *
PlatformPS4       -0.567574   0.288393  -1.968            0.049061 *
PlatformPSP       -0.004297   0.223161  -0.019            0.984639
PlatformPSV       -0.394330   0.309315  -1.275            0.202362
PlatformWii        0.391461   0.209162   1.872            0.061267 .
PlatformWiiU      -0.452494   0.377504  -1.199            0.230665
PlatformX         -0.235119   0.304770  -0.771            0.440433
PlatformX360      -0.224743   0.214586  -1.047            0.294947
PlatformXOne      -1.743726   0.388584  -4.487   0.0000072104475415 ***
Avg_Score          0.103891   0.023644   4.394   0.0000111345932325 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 9132.2  on 8221  degrees of freedom
Residual deviance: 4370.0  on 8194  degrees of freedom
AIC: 4426

Number of Fisher Scoring iterations: 8
```

Multiple Linear Regression

- Target Variable – Global_Sales
- Prediction – To predict the Total Global Sales based on the model.
- Independent Variables – Platform, Genre, and Avg_Score

```
> videogamepred <- data.frame(Platform="PC", Genre = 'Strategy', Avg_Score = 7.5)
> sale_pred <- predict(videogamelm, newdata = videogamepred, type="response")
> sale_pred
       1
0.2105195
> videogamepred <- data.frame(Platform="PS4", Genre = 'Role-Playing', Avg_Score = 9)
> sale_pred <- predict(videogamelm, newdata = videogamepred, type="response")
> sale_pred
       1
1.179562
```

```
Call:
lm(formula = Global_Sales ~ Platform + Genre + Avg_Score, data = train.df)

Residuals:
   Min     1Q Median     3Q    Max
-1.331 -0.444 -0.196  0.103 34.433

Coefficients:
                Estimate Std. Error t value             Pr(>|t|)
(Intercept)      -0.27937    0.08533  -3.274              0.00107 **
PlatformDS       -0.10363    0.07935  -1.306              0.19161
PlatformGBA      -0.32831    0.14110  -2.327              0.02000 *
PlatformGC       -0.40305    0.15125  -2.665              0.00772 **
PlatformPC       -0.46504    0.08935  -5.205       0.000000198916 ***
PlatformPS2      -0.15274    0.08661  -1.763              0.07786 .
PlatformPS3       0.11247    0.08379   1.342              0.17957
PlatformPS4       0.18493    0.10595   1.745              0.08094 .
PlatformPSP      -0.22912    0.08514  -2.691              0.00714 **
PlatformPSV      -0.30395    0.10340  -2.940              0.00330 **
PlatformWii       0.13902    0.08431   1.649              0.09921 .
PlatformWiiU     -0.12871    0.14758  -0.872              0.38316
PlatformX        -0.50907    0.12233  -4.161       0.000031951179 ***
PlatformX360      0.13191    0.08507   1.551              0.12104
PlatformXOne      0.01352    0.12379   0.109              0.91304
GenreAdventure   -0.10256    0.05800  -1.768              0.07708 .
GenreFighting    -0.03335    0.08095  -0.412              0.68037
GenreMisc         0.02516    0.05386   0.467              0.64041
GenrePlatform     0.21143    0.08306   2.546              0.01093 *
GenrePuzzle      -0.09207    0.08792  -1.047              0.29504
GenreRacing       0.15056    0.07085   2.125              0.03362 *
GenreRole-Playing  0.04453    0.05791   0.769              0.44191
GenreShooter      0.41485    0.06319   6.565       0.000000000055 ***
GenreSimulation   0.04477    0.07034   0.637              0.52446
GenreSports       0.07034    0.05362   1.312              0.18957
GenreStrategy    -0.06963    0.08317  -0.837              0.40252
Avg_Score         0.13661    0.00758  18.021 < 0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.342 on 8195 degrees of freedom
Multiple R-squared:  0.07734,   Adjusted R-squared:  0.07441
F-statistic: 26.42 on 26 and 8195 DF,  p-value: < 0.00000000000000022

> videogame.df.pred <- predict(videogamelm, valid.df)
> accuracy(videogame.df.pred, valid.df$Global_Sales)
                 ME      RMSE       MAE       MPE     MAPE
Test set 0.004341261 1.781781 0.5250464 -359.4439 645.5337
```

Final Model

- Target Variable – Success
- Prediction –
  - Probability of Success of a title
  - Platform and genre of a title
- Model – Combination of Logistic Regression, & Decision Tree

Why?

- Logistic Regression provides insights about the important categorical variables
- Decision Tree, with the highest accuracy, predicts the success probability of a title

Limitations

- Predictions are based on some outdated Platforms e.g Wii.
- The model does not take into consideration the number of active users (Data Unavailable)
- Does not consider new user sign up rate (Data Unavailable)

Improvements

- Aggregating platforms with generations e.g. PS3,PS4, to be considered as PS
- Obtaining Active User statistics
- Obtaining more recent data (From 2017 onwards)

Recommendation

```
> table(success.predlogistic, valid.df$Genre)

success.predlogistic Action Adventure Fighting Misc Platform Puzzle Racing Role-Playing Shooter Simulation
                   1    183        22       33   72       34     12     39           86      74         34

success.predlogistic Sports Strategy
                   1     87       10
> table(success.predlogistic, valid.df$Platform)

success.predlogistic 3DS  DS GBA  GC  PC PS2 PS3 PS4 PSP PSV Wii WiiU   X X360 XOne
                   1  30  93   7   5  25  79 119  35  40   7  95   15   8  111   17
```

Most Successful Genre:

- Action
- Sports
- Role-Playing

Most Successful Platform:

- Playstation
- Nintendo
- Xbox

Recommendation

- Target PlayStation as a platform for the next game as per Logistic Regression
- Develop the game belonging to Action or Sports genre as per Logistic Regression
- The target Avg_Rating for the title should be >=7.3 as per Logistic Regression

Thank you