2. Classification on \_2016\_>game -> dataset

---

title: "classification"

author: "RStudio"

Refrence:"https://rpubs.com/vh42720/vgsales,https://www.kaggle.com/anupambera/video-game-sales-analysis"

date: "28/06/2021"

output: html\_document

---

```{r setup, include=FALSE}

knitr::opts\_chunk$set(echo = TRUE)

```

```{r}

library(tidyverse)

library(ggplot2)

library(ggthemes)

library(viridis)

library(lubridate)

library(wordcloud)

library(qdap)

library(tm)

library(ngram)

library(RColorBrewer)

library(gridExtra)

library(knitr)

```

```{r}

game <- tibble::as\_tibble(read.csv("C:/Users/priya/Desktop/Game Research/Video\_Games\_Sales\_as\_at\_22\_Dec\_2016.csv", stringsAsFactors = FALSE))

theme\_set(theme\_tufte())

```

```{r}

str(game)

```

```{r}

colSums(is.na(game))

#There are quite a lot of missing values in score but they appear in Sales which means a lot of games are neglected by the rating community. It is worth our time to take a closer look at these titles in missing values analysis.

```

## General Histograms

```{r pressure, echo=FALSE}

game %>%

group\_by(Year\_of\_Release) %>%

summarize(Sales = sum(Global\_Sales, na.rm = TRUE)) %>%

ggplot(aes(x = Year\_of\_Release, y = Sales)) +

geom\_col(fill = "navyblue") +

theme(axis.text.x = element\_text(angle = 90)) +

labs(title = "Global Sales Histograms", x = "Year", y = "Sales (units)")

```

```{r}

game %>%

group\_by(Year\_of\_Release) %>%

summarize(Number\_of\_Games = n()) %>%

ggplot(aes(x = Year\_of\_Release, y = Number\_of\_Games)) +

geom\_col(fill = "magenta4") +

theme(axis.text.x = element\_text(angle = 90)) +

labs(title = "Games released per Year", x = "Year", y = "Sales (units)")

```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

# Top Global Sales each year

```{r}

game %>%

group\_by(Year\_of\_Release, Publisher) %>%

summarize(Sales = sum(Global\_Sales)) %>%

top\_n(n = 1) %>%

kable()

```

Missing Values Analysis

All missing values are in Critic Score, Critic Count and User Count. However, Developer also has an empty string for these missing values. Many of these titles are quite famous such as Super Mario Bros… thus perhaps they were released before the rating website created. A simple plot would suffice.

```{r}

game$Year\_of\_Release <- as.integer(game$Year\_of\_Release)

game %>%

filter(is.na(Critic\_Score)) %>%

ggplot(aes(x = Year\_of\_Release, y = Genre)) +

geom\_jitter(alpha = 0.2, pch = 21, col = "darkslategrey") +

labs(x = "Year of Release", title = "Missing Score by Genre and Year")

```

It seems our intuition is wrong; most missing scores are from 2005 and up. Nevertheless, there are a few interesting points show up in the plot:

Action, Adventure, Sport, Simulation and Misc have the most missing values. This trend correlated heavily with Genre distribution so it is not a surprise.

There are 2 games from 1993 that have no Genres listed. They also have no name, created by Acclaim Entertainment for GEN gaming system. Their sales combine to a total of less than 2.5 millions. These will be excluded from our dataset.

Imagine: Makeup Artist is/was/will be released in 2020 which proves time travel! Joking aside, it has already presale for $ 290.000 total.

Lastly, there are only 3 games from 2017 and we should also exclude them

```{r}

index <- which(game$Genre == "" | game$Year\_of\_Release %in% c(2020, 2017))

game <- game[-index,]

## Creating System Variable

pc <- c("PC")

console <- c("Wii", "NES", "X360", "PS3", "PS2", "SNES", "PS4", "N64", "PS", "XB", "2600",

"XOne", "WiiU", "GC", "GEN", "DC", "SAT", "SCD", "NG", "TG16", "3DO", "GG",

"PCFX")

handheld <- c("GB", "DS", "GBA", "3DS", "PSP", "PSV", "WS")

game <- game %>%

mutate(System = ifelse(Platform %in% pc, "pc",

ifelse(Platform %in% console, "console", "handheld")))

```

```{r}

max\_mis <- game %>% filter(is.na(Critic\_Score)) %>%

group\_by(System) %>% filter(Global\_Sales == max(Global\_Sales))

game %>%

filter(is.na(Critic\_Score)) %>%

ggplot(aes(x = Platform, y = Global\_Sales, col = System)) +

geom\_point(alpha = 0.5, pch = 21) +

geom\_text(data = max\_mis, label = max\_mis$Name, vjust = "inward", hjust = "inward") +

coord\_flip() +

labs(title = "Missing Score sorted by Global Sales and Platform")

```

Even with a missing score, these games still generate a healthy amount of revenue with Super Mario Bros earned over 40 millions in sales. The data shows no trend that any system got an edge in getting scored.

To conclude the missing values portion of this report, lets take a look at the wordcloud.

```{r}

clean <- function(x){

x <- replace\_contraction(x)

x <- replace\_ordinal(x)

x <- removePunctuation(x)

x <- tolower(x)

x <- removeWords(x, c(stopwords("en"), "game"))

}

missing <- filter(game, is.na(Critic\_Score))

titles <- concatenate(missing$Name)

#titles <- clean(titles)

unigram <- ngram(titles, n = 1)

freq <- tbl\_df(get.phrasetable(unigram))

## Creating wordcloud

wordcloud(freq$ngrams, freq$freq, random.order = FALSE, rot.per = 0.35,

color=brewer.pal(8, "Dark2"), max.words = 200)

```

There are a lot of sequels among these missing scores. Other words are quite typical such as pro, world, soccer, portable, battle, gundam, edition, championship…

Sales vs Devs, Publisher and Platform

First, lets pay tribute to the people who make and distribute our games.

```{r}

game %>%

select(Publisher, Developer) %>%

sapply(function(x) {length(unique(x))})

```

```{r}

top20 <- head(names(sort(table(game$Publisher), decreasing = TRUE)),20)

p1 <- game %>%

filter(Publisher %in% top20) %>%

mutate(Publisher = factor(Publisher, levels = rev(top20))) %>%

group\_by(Publisher) %>%

summarize(Number\_of\_Devs = n\_distinct(Developer)) %>%

top\_n(20, wt = Number\_of\_Devs) %>%

ggplot(aes(x = Publisher, y = Number\_of\_Devs)) +

geom\_col(fill = "darkslategrey", alpha = 0.5) +

coord\_flip()

p2 <- game %>%

filter(Publisher %in% top20) %>%

mutate(Publisher = factor(Publisher, levels = rev(top20))) %>%

group\_by(Publisher) %>%

summarize(Number\_of\_Games = n()) %>%

top\_n(20, wt = Number\_of\_Games) %>%

ggplot(aes(x = Publisher, y = Number\_of\_Games)) +

geom\_col(fill = "firebrick4", alpha = 0.5) +

coord\_flip() +

theme(axis.title.y = element\_blank(), axis.text.y = element\_blank())

grid.arrange(p1,p2, ncol = 2, top = "Distribution of games and devs for top 20 Publishers")

```

There are a lot more Developers than Publishers. In fact, most big Publishers have over 100 Developers working for them. Notice the Unknown Publisher is in the top 20 also. Unknown here means the collection of unrecorded publisher and not the name “Unknown”.

The more interesting point is the massive amount of titles big Publishers have compare to the rest. Even with only the top 20 we can clearly see the big difference in distributive power. How about the sale?

###### Time series ########

```{r}

toppub <- game %>%

filter(Publisher %in% c("Electronic Arts", "Activision", "Ubisoft", "Namco Bandai Games", "Nintendo")) %>%

group\_by(Publisher, Year\_of\_Release) %>%

summarize(total = sum(Global\_Sales))

game %>%

group\_by(Publisher, Year\_of\_Release) %>%

summarize(total = sum(Global\_Sales)) %>%

ggplot(aes(x = Year\_of\_Release, y = total)) +

geom\_point(alpha = 0.5, pch = 21) +

geom\_point(data = toppub, aes(col = Publisher), size = 1.5) +

geom\_line(data = toppub, aes(col = Publisher), size = 1.3) +

scale\_color\_viridis(discrete = TRUE) +

theme(legend.position = c(0.2, 0.7)) +

labs(title = "Time series of Global Sales by Publishers", y = "Sales (in million units)")

```

The sales distribution is amazing! The top 5 Publishers take home multiple times the sales of the rest combined. While it is expected given the distribution of game titles and developers, the plot shows how hard it is to compete with these big buys in the industry.

EA and Ubisoft joined late but started to dominate the market very quickly. Note that EA overtaked most of the competition by 2010. However, their practice of buying out their competition is not very popular with the fans.

### EA acquisitions list ###

# EA reddit breakdown #

```{r}

toppub <- game %>%

filter(Publisher %in% c("Electronic Arts", "Activision", "Ubisoft", "Namco Bandai Games", "Nintendo")) %>%

gather("Region", "Sales", ends\_with("Sales")) %>%

group\_by(Publisher, Year\_of\_Release, Region) %>%

filter(Region != "Global\_Sales") %>%

summarize(total = sum(Sales))

game %>%

gather("Region", "Sales", ends\_with("Sales")) %>%

group\_by(Publisher, Year\_of\_Release, Region) %>%

filter(Region != "Global\_Sales") %>%

summarize(total = sum(Sales)) %>%

ggplot(aes(x = Year\_of\_Release, y = total)) +

geom\_point(alpha = 0.5, pch = 21) +

facet\_wrap(~Region, scales = "free\_y") +

geom\_point(data = toppub, aes(col = Publisher), size = 0.5) +

geom\_line(data = toppub, aes(col = Publisher), size = 1.1) +

scale\_color\_viridis(discrete = TRUE) +

theme(legend.position = "top", axis.title.x = element\_blank()) +

labs(title = "Time series of Global Sales by Publishers and Region", y = "Sales (in million units)")

```

When sorted through regions, sales distribution is the same for NA and US. However, Nintendo and Manco Bandai Games are the only big players in Japan market.

Moreover, NA an EU markets generates much more sales than Japan and others - population matters here.

Notice there is a downward trend after 2010 in sales across the industry.

Reproducing the above decline using different dataset is reccomended to reach any conclusion.

It is not a gaming analysis without system war. Three contenders are: console, PC and handheld.

```{R}

game %>%

group\_by(System, Year\_of\_Release) %>%

summarize(total = sum(Global\_Sales, na.rm = TRUE)) %>%

ggplot(aes(x = Year\_of\_Release, y = total, fill = System)) +

geom\_col(position = "stack") +

labs(y = "Sales (in million $)", title = "Global Sales by System and time") +

scale\_fill\_viridis(discrete = TRUE)

```

This is a dangerous statistic. While the data clearly shows PC games sales is practically nothing compared to the other two, remember most revenue of PC games come from MOBA, RPG, … which are free-to-play!

# PC generates more revenue

Lets take a closer look at the console markets where there are three big players: xbox, playstation, nintendo and others. The classification only include the big names and gloat over many. However, it should still give us a fair observation

```{r}

xbox <- c("X360", "XB", "XOne")

nintendo <- c("Wii", "WiiU", "N64", "GC", "NES")

playstation <- c("PS", "PS2", "PS3", "PS4")

game %>%

filter(System == "console") %>%

mutate(console\_type = ifelse(Platform %in% xbox, "xbox",

ifelse(Platform %in% nintendo, "nintendo",

ifelse(Platform %in% playstation, "playstation",

"others")))) %>%

group\_by(console\_type, Year\_of\_Release) %>%

summarize(total = sum(Global\_Sales)) %>%

ggplot(aes(x = Year\_of\_Release, fill = console\_type)) +

geom\_density(position = "fill") +

labs(title = "Distribution of game titles sorted by console type") +

scale\_fill\_viridis(discrete = TRUE)

```

```{r}

game %>%

filter(System == "console", Year\_of\_Release <= 2010) %>%

mutate(console\_type = ifelse(Platform %in% xbox, "xbox",

ifelse(Platform %in% nintendo, "nintendo",

ifelse(Platform %in% playstation, "playstation",

"others")))) %>%

group\_by(console\_type, Year\_of\_Release) %>%

summarize(total = sum(Global\_Sales)) %>%

ggplot(aes(x = Year\_of\_Release, y = total, col = console\_type)) +

geom\_point(alpha = 0.2) +

geom\_line(alpha = 0.2) +

stat\_smooth(method = "lm", se = FALSE) +

theme(legend.position = c(0.2, 0.7)) +

labs(title = "Console sales before 2010 across console platform", y = "Sales (in units)") +

scale\_color\_viridis(discrete = TRUE)

```

```{r}

game %>%

filter(System == "console", Year\_of\_Release > 2010) %>%

mutate(console\_type = ifelse(Platform %in% xbox, "xbox",

ifelse(Platform %in% nintendo, "nintendo",

ifelse(Platform %in% playstation, "playstation",

"others")))) %>%

group\_by(console\_type, Year\_of\_Release) %>%

summarize(total = sum(Global\_Sales)) %>%

ggplot(aes(x = Year\_of\_Release, y = total, col = console\_type)) +

geom\_point(alpha = 0.2) +

geom\_line(alpha = 0.2) +

stat\_smooth(method = "lm", se = FALSE) +

labs(title = "Console sales after 2010 across console platform", y = "Sales (in units)") +

scale\_color\_viridis(discrete = TRUE)

```

Few observations:

Nintendo was the only big player before 1995 and thus they had a complete control over the market

After 1995, Xbox and playstation gained the majority of market shares

There are only 3 big players in console in the last 20 years, all others are going extinct.

Again we see a sharp decline in sales after 2010

## Sales vs Genre and Rating ##

If you are not familiar with games rating, here is the short version. Note that we have an empty string which means missing value for one of the rating category

AO - Adult only 18+ M - Mature Only 17+ T - Teen E10+ - Everyone 10+ E - Everyone EC - Early Childhood RP - Rating Pending

```{r}

game %>%

ggplot(aes(x = Rating, y = Genre, col = Genre)) +

geom\_jitter(alpha = 0.4, pch = 21) +

theme(legend.position = "none") +

scale\_color\_viridis(discrete = TRUE)

```

There are many games without rating. In contrast, most Publisher avoids K-A, EC and AO rating since the playerbase is too specific. Genre is not affected by time as much.

```{r}

game %>%

group\_by(Genre, Name) %>%

summarise(Sales = sum(Global\_Sales)) %>%

top\_n(n = 1) %>%

kable()

```

Lastly, lets take a look at how different regions favor different genre.

```{r}

game %>%

gather("Region", "Value", c("NA\_Sales", "EU\_Sales", "JP\_Sales", "Other\_Sales")) %>%

group\_by(Region, Genre) %>%

summarize(Sales = sum(Value)) %>%

top\_n(n = 3) %>%

ggplot(aes(x = Region, y = Sales, group = Region, fill = Genre)) +

geom\_col(position = "stack") +

scale\_fill\_viridis(discrete = TRUE) +

labs(title = "Top Genre by Sales per Region")

```

```{r}

names <- concatenate(vgsales$Name)

names <- clean(names)

unigram <- ngram(names, n = 1)

freqs <- get.phrasetable(unigram)

wordcloud(freqs$ngrams, freqs$freq, random.order = FALSE, rot.per = 0.35,

color=brewer.pal(8, "Dark2"), max.words = 100)

```

##

Video Game is an electronic game that is played on electronic medium devices such as personal computer, TV screen, gaming console or mobile phone. Some time the Video Game industry is called the interactive entertainment industry. The input device used for games, the game controller, varies across platforms. Common controller includes game pad, joysticks, mouse, keyboard, the touchscreens of mobile devices and buttons. Players typically view the game on a video screen or television and there are often game sounds from loudspeakers. Video Game development has a long history since 1970’s and in recent past with the revolution of the smartphones and tablets introduced new categories of video games such as mobile and social games. Developers introduced various technology and methodology in the computing system to popularize and make the video game more interesting and interactive such as “virtual reality”.

The motivation of this project is to visualize the data set and practice exploratory analysis of data set. For the better understanding I have analyzed the data by some histograms and plot, which will help us to know the trend of the industry. I used some statistical methods to fit the data set and predict the sales.

```{r}

library(e1071)

library(tree)

library(MASS)

game=na.omit(game)

attach(game)

game$Name=as.factor(as.character(game$Name))

game$Platform=as.factor(as.character(game$Platform))

game$Year\_of\_Release=as.numeric(as.character(game$Year\_of\_Release))

game$Genre=as.factor(as.character(game$Genre))

game$Publisher=as.factor(as.character(game$Publisher))

max(game$Year\_of\_Release,na.rm=T)

#Histogram of frequency of the game by year

hist(game$Year\_of\_Release,col = "blue",xlab = "Year",ylab = "Frequency of the game", main = "Histogram of frequency of the game by year")

```

```{R}

#Histogram of Global sales of the game by year

aggregate\_revenue=aggregate(Global\_Sales~Year\_of\_Release,game,sum)

plot(aggregate\_revenue,type='h',xlab="Year",ylab="Global Sales",col = "green", lwd = 8, main = "Global Sales per year")

```

Observation: Total revenue was increasing till 2008 after that it is in declined. Maximum revenue was in 2008 and 2009.

```{r}

#Histogram of Global sales of the game by genre

revenue\_by\_Genre=aggregate(Global\_Sales~Genre,game,sum)

arrange\_by\_Genre=arrange(revenue\_by\_Genre,desc(Global\_Sales))

arrange\_by\_Genre$Genre = factor(arrange\_by\_Genre$Genre, levels = arrange\_by\_Genre$Genre)

ggplot(arrange\_by\_Genre,aes(Genre,Global\_Sales)) +

geom\_bar(fill="blue",stat = "identity") +

ggtitle("Video Game - Global Sales by Genre")

```

Action type of game is more popular and it generates the maximum revenue. Second highest sales generating game is Sports type followed by Shooter, Role-Playing, Platform, Misc, Racing, Fighting, Simulation, Puzzle, Adventure and Strategy.

```{r}

#Histogram of top 10 Publisher by revenue

revenue\_of\_Publisher=aggregate(Global\_Sales~Publisher,game,sum)

arrange\_Revenue\_of\_Publisher\_by\_Global\_Sales=arrange(revenue\_of\_Publisher,desc(Global\_Sales))

top\_10=arrange\_Revenue\_of\_Publisher\_by\_Global\_Sales[1:10,]

#plot(factor(top\_10$Publisher),top\_10$Global\_Sales,type='h',lwd = 8,xlab="Publisher",ylab="Global Sales",col = "red",main = "Top 10 Publishers by Revenue")

ggplot(top\_10,aes(Publisher,Global\_Sales, fill=Publisher))+

geom\_bar(stat = "identity")+

ggtitle("Top 10 Publisher by Revenue") +

theme(legend.position = "top")

```

Observation: Nintendo is a top publisher followed by Electronic Arts, Activision, Sony Computer Entertainment, Ubisoft, Take-Two Interactive, THQ, Konami Digital Entertainment, Sega and Namco Bandai Games

```{r}

#Top 10 game by Revenue in each year

arrange\_by\_Year\_of\_Release\_and\_Name=game %>%

group\_by(Year\_of\_Release, Name) %>%

summarize(Toal\_Global\_Sales = sum(Global\_Sales)) %>%

arrange(desc(Toal\_Global\_Sales)) %>%

top\_n(1)

top10\_revenue\_by\_Year\_of\_Release\_and\_Name=arrange\_by\_Year\_of\_Release\_and\_Name[1:10,]

ggplot(top10\_revenue\_by\_Year\_of\_Release\_and\_Name,aes(x=as.factor(Year\_of\_Release),Toal\_Global\_Sales,fill=Name))+

geom\_bar(stat = "identity")+

ggtitle("Top 10 Games by Revenue each year") +

theme(legend.position = "top")

```

Wii Sports is the top performer on 2006 and it generates the revenue globally 82.74 million dollars followed by Super Mario Bros (In 1985 – 40.24m)andGrandTheftAutoV(In2013− 37.78 m).

```{r}

#Correlation of the sales Factor

num\_Sales=game[,c("NA\_Sales","EU\_Sales","JP\_Sales","Other\_Sales","Global\_Sales")]

cor(num\_Sales)

```

From the correlation table it is observed that the NA\_Sales (0.94), EU\_Sales (0.9) and Other\_Sales (0.75) are highly positive correlated with the Global\_Sales. Although Global\_Sales is correlated with the all Sales regions.

```{r}

# Prediction : Support Vector Machines

train =( Year\_of\_Release <= 2012)

game.train=game[train,]

game.test = game[!train,]

# Linear classification

y.train=ifelse(game.train$Global\_Sales>10.0,1,-1)

dat=data.frame(x=game.train$NA\_Sales+game.train$EU\_Sales, y=as.factor(y.train))

svmfit=svm(y~., data=dat, kernel="linear", cost=10,scale=FALSE)

#summary(svmfit)

table(Model=svmfit$fitted , Truth=dat$y)

cat("Model Error = ", mean(svmfit$fitted!=dat$y)\*100,"%")

y.test=ifelse(game.test$Global\_Sales>10.0,1,-1)

dat.te=data.frame(x=game.test$NA\_Sales+game.test$EU\_Sales, y=as.factor(y.test))

pred.te=predict(svmfit, newdata=dat.te)

table(Predict=pred.te, Truth=dat.te$y)

cat("Prediction Error = ", mean(pred.te!=dat.te$y)\*100,"%")

```

SVM Linear - Prediction error of Global Sales will more than 10 million dollar, using North American Sales and European Sales data is 0.096%

```{r}

# Radial classification

svmfit=svm(y~., data=dat, kernel="radial", cost=10, gamma=1)

table(Model=svmfit$fitted , Truth=dat$y)

cat("Model Error = ", mean(svmfit$fitted!=dat$y)\*100,"%")

y.test=ifelse(game.test$Global\_Sales>10.0,1,-1)

dat.te=data.frame(x=game.test$NA\_Sales+game.test$EU\_Sales, y=as.factor(y.test))

pred.te=predict(svmfit, newdata=dat.te)

table(predict=pred.te, truth=dat.te$y)

cat("Prediction Error = ", mean(pred.te!=dat.te$y)\*100,"%")

```

SVM Radial Classification - Prediction error of Global Sales will more than 10 million dollar, using North American Sales and European Sales data is 0.096%

```{r}

#Prediction through Decision Trees

num\_fact=game[,c("NA\_Sales","EU\_Sales","Global\_Sales")]

High=ifelse(Global\_Sales <=10.0,"No","Yes")

dat =data.frame(num\_fact,High)

tree.dat=tree(as.factor(High)~.-Global\_Sales, dat,subset=train)

summary(tree.dat)

plot(tree.dat)

text(tree.dat ,pretty =0)

dat.test = game[!train,]

High.test=High[!train]

tree.pred=predict(tree.dat,dat.test,type="class")

table(Predict=tree.pred ,Truth=High.test)

cat("Prediction Error = ", mean(tree.pred!=High.test)\*100,"%")

```

Decision Tree - Prediction error of Global Sales will more than 10 million dollar, using North American Sales and European Sales data is 0.19%.

```{r}

#Prediction through Linear Regression

train =( Year\_of\_Release <= 2012)

num\_fact=game[,c("NA\_Sales","EU\_Sales","Global\_Sales")]

High=ifelse(Global\_Sales <=10.0,"No","Yes")

dat =data.frame(num\_fact,High)

glm.fit=glm(as.factor(High)~.-Global\_Sales ,dat,subset=train,family=binomial)

summary(glm.fit)

coef(glm.fit)

summary(glm.fit)$coef

dat.test = game[!train,]

High.test=High[!train]

glm.prob=predict(glm.fit,dat.test,type="response")

glm.pred=rep("No",dim(dat.test)[1])

glm.pred[glm.prob >.5]="Yes"

table(Predict=glm.pred ,Truth=High.test)

cat("Prediction Error = ", mean(glm.pred!=High.test)\*100,"%")

```

Linear Regression by Generalized Linear Models - Prediction error of Global Sales will more than 10 million dollar, using North American Sales and European Sales data is 0.19%.

```{r}

#Linear Discriminant Analysis

lda.fit=lda(as.factor(High)~.-Global\_Sales ,dat,subset=train)

summary(lda.fit)

plot(lda.fit)

dat.test = game[!train,]

High.test=High[!train]

lda.pred=predict(lda.fit,dat.test)

names(lda.pred)

lda.class=lda.pred$class

table(Predict=lda.class ,Truth=High.test)

cat("Prediction Error = ", mean(lda.class!=High.test)\*100,"%")

```

Linear Regression by Linear Discriminant Analysis - Prediction error of Global Sales will more than 10 million dollar, using North American Sales and European Sales data is 0.14%.

## Final Notes ##

Game sales peaked in 2008 and has been decreasing ever since. However, the data is not a representation of overall gaming scene. Modern games focus on free to play platform, dominated by PC games.

The data is flawed since it only takes into account game sales where most revenue from games comes from microtransactions. Again, we did not have the data to show PC games on the rise.

EA, Nintendo, Sony are the last few big companies can manage to survive

Most bad games die out instantly in the last few years which signify the rise in standard of games.

The analysis focuses on consoles due to the data structure.