B.Tech, Sem III
Session : Aug-Dec, 2019

UE18CS203 – INTRODUCTION TO DATA SCIENCE

REPORT
ON
EXPLORATORY ANALYSIS ON
VIDEO GAME SALES

SECTION : A

#	SRN	Name	Contact No.	Email ID	Sign
1.	PES1201801041	Harichandana Magapu	8105642319	cmagapu@gmail.com	
2.	PES1201800883	Avinash Ratnam	9980673619	avinashkratnam@gmail.com	

ABOUT THE DATA SET

The data set given to us contained an extensive report on the video game sales across North America, Europe, Japan and the rest of the world. This data set also helps give an insight on how the sales have changed from year to year, the trends that publishers used to develop certain games for maximum profit and the popularity of certain genres over the years. The rows included are:

- <u>Name</u>: Title of the game.
- <u>Platform</u>: The medium on which the game was released. (Egs- PC, Xbox One, PS3 etc.)
- <u>Year</u>: The year of release (between 1980 - 2020)
- <u>Genre</u>: The genre of the game. (Egs – Action, Puzzle etc.)
- <u>Publisher</u>: The developers of the game (Egs – EA, Ubisoft etc.).
- <u>North America Sales</u>: The total number of units sold in the continent of North America for a particular game (in millions).
- <u>Europe Sales</u>: The total number of units sold in the continent of Europe for a particular game (in millions).
- <u>Japan Sales</u>: The total number of units sold in the country of Japan for a particular game (in millions). Japan has been added separately since it accounts for approximately 25% of the total Asian sales revenue.
- <u>Other Sales</u>: The total number of units sold in the rest of the world for a particular game (in millions).
- <u>Global Sales</u>: The total number of units sold for a game across the world (inclusive of NA, EU and JP sales).

The data set has a total of 16,598 elements.

ABSTRACT

The video game industry as of 2018 is worth $135 Billion beating out the second place TV Industry by approximately $21 Billion, and the industry is only growing. A steady rate of 10.6% increase in sales figures per year as compared to a decline of 8% to the second place TV revenue. Other reasons for this rapid growth could be credited to the rise in eSports, Artificial Reality gaming and plenty more improvements in the industry. With this dataset, we get an insight on how companies perform in different regions, how they specifically focus on a certain game genre to maximize sales.
But with an industry this profitable and rapid, there are bound to be questions on whether companies are actually making profits acceptable enough to sustain ? Which genres are more profitable, does the sales region play a role in the sales figures ? Which genre sells better on which console ? Which is the most profitable genre? Which companies are creating the most revenue and which aren't even considered anymore ?
Hence we used several visual representation such as Bar Charts, Pie Charts, Scatter Plots etc. to get a clear and direct answer to all our queries.

EXPLORATORY ANALYSIS

The dataset provided did have some missing values regarding the 'Publisher' and the 'Year' columns. Since these unknown values come from categorical data, dropping these rows did not affect the outcomes drastically.

Due to the large number of values we were working with however, we created a separate entity for the Big 3 publishers (EA, Sony and Nintendo) to help dig deeper into their statistics.
The total number of elements we ended up working with was 16,291 (originally 16,598).

Before cleaning the dataset

	Rank	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
count	16598.000000	16327.000000	16598.000000	16598.000000	16598.000000	16598.000000	16598.000000
mean	8300.605254	2006.406443	0.264667	0.146652	0.077782	0.048063	0.537441
std	4791.853933	5.828981	0.816683	0.505351	0.309291	0.188588	1.555028
min	1.000000	1980.000000	0.000000	0.000000	0.000000	0.000000	0.010000
25%	4151.250000	2003.000000	0.000000	0.000000	0.000000	0.000000	0.060000
50%	8300.500000	2007.000000	0.080000	0.020000	0.000000	0.010000	0.170000
75%	12449.750000	2010.000000	0.240000	0.110000	0.040000	0.040000	0.470000
max	16600.000000	2020.000000	41.490000	29.020000	10.220000	10.570000	82.740000

After cleaning the dataset

|	Rank	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
count	16291.000000	16291.000000	16291.000000	16291.000000	16291.000000	16291.000000	16291.000000
mean	8290.190228	2006.405561	0.265647	0.147731	0.078833	0.048426	0.540910
std	4792.654450	5.832412	0.822432	0.509303	0.311879	0.190083	1.567345
min	1.000000	1980.000000	0.000000	0.000000	0.000000	0.000000	0.010000
25%	4132.500000	2003.000000	0.000000	0.000000	0.000000	0.000000	0.060000
50%	8292.000000	2007.000000	0.080000	0.020000	0.000000	0.010000	0.170000
75%	12439.500000	2010.000000	0.240000	0.110000	0.040000	0.040000	0.480000
max	16600.000000	2020.000000	41.490000	29.020000	10.220000	10.570000	82.740000

Firstly, we wanted to see how the gaming industry has improved since its inception in 1980. For this we plotted a line graph between the 'Global Sales' and the 'Year'.
As expected from the graph below we see a steady flow in the sales figures between the 1980 and 90's and a sharp rise at the turn of the century in the early 2000's.
We can deduce that 2009-10 was the peak of the gaming industry with the maximum sales.

A useful follow up graph would be to analyze which genre sells the best.

Total Genre Sales
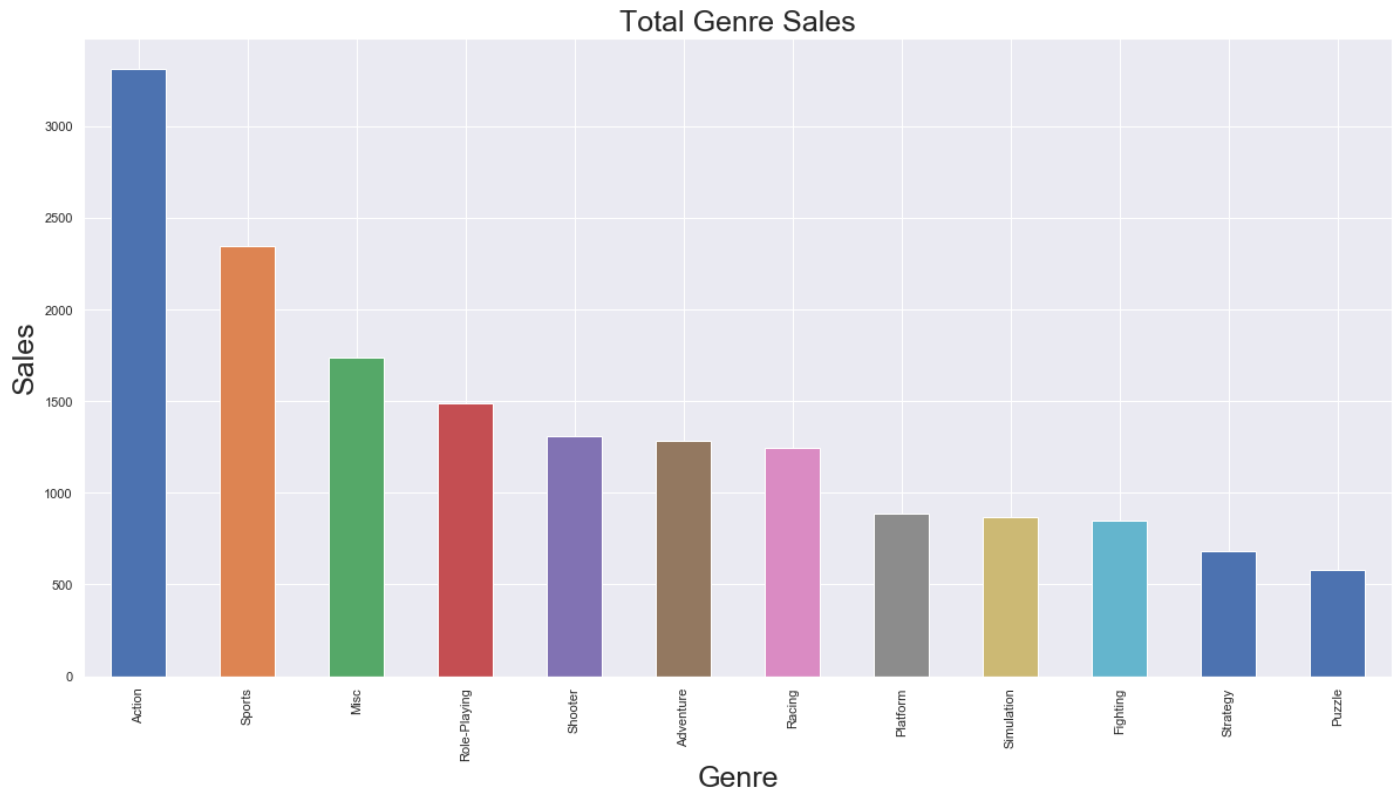
Clearly, Action games is the dominant genre, but which region is responsible for these massive sales ?

From the below pie chart, we see North America contributes to nearly 50% of the total sales of Action games. This can be credited to the fact that all major action game publishers are all based in North America. Other notable factors include popularity of eSports in North America and general affordability of such AAA Games. (AAA Games are games that were developed under the highest development budget of that current time and are highest promoted.



Action Game Sales by Region
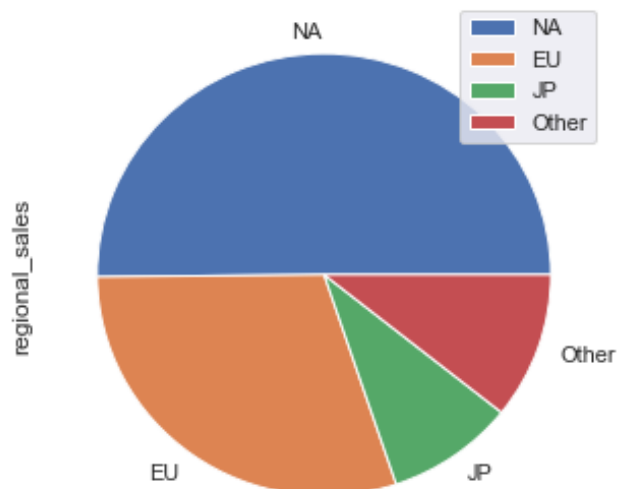
A fair assumption would be that consoles with better sales figures with respect to games sold would sell better as well. Let's see if that is true.

Here we see that the DS and PS2 had very similar sales figures, but in terms of their console sales ?
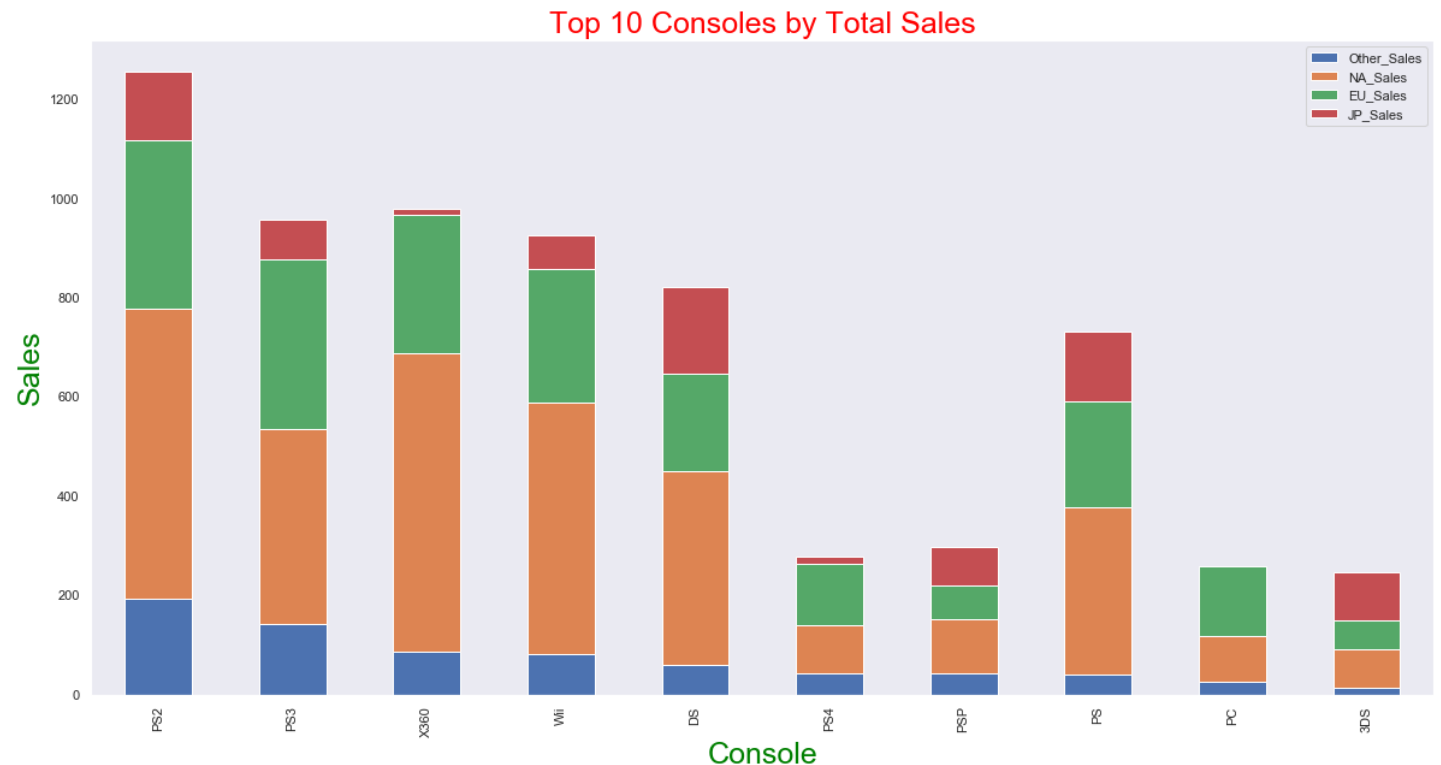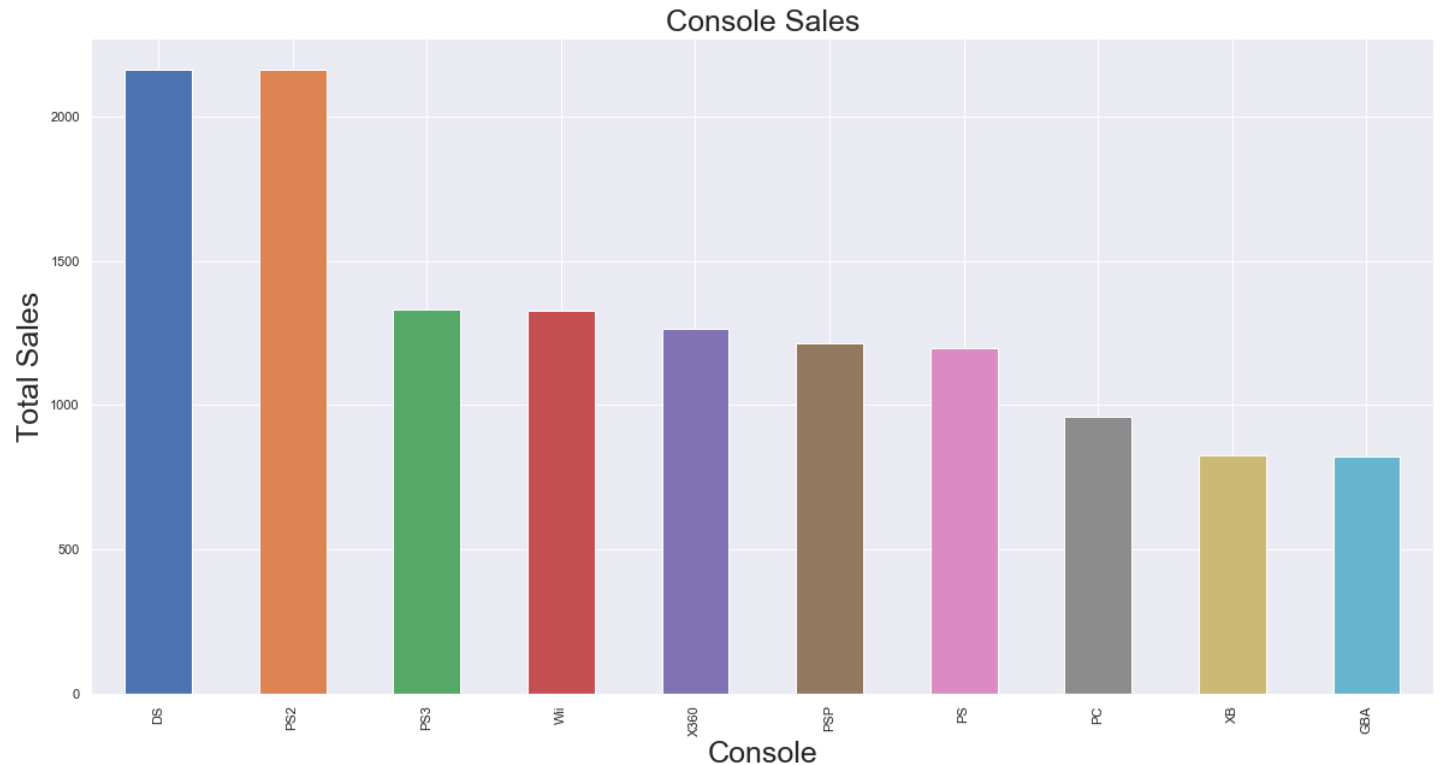

The PS2 completely overshadows the DS in actual console sales, with North American sales of the DS being more than the North America and Rest of World Sales of the DS combined.

There are almost 579 publishers listed in the dataset. Which one's are the most popular ? Is it safe to assume that the region from which the publisher is based is where it gains maximum sales ?
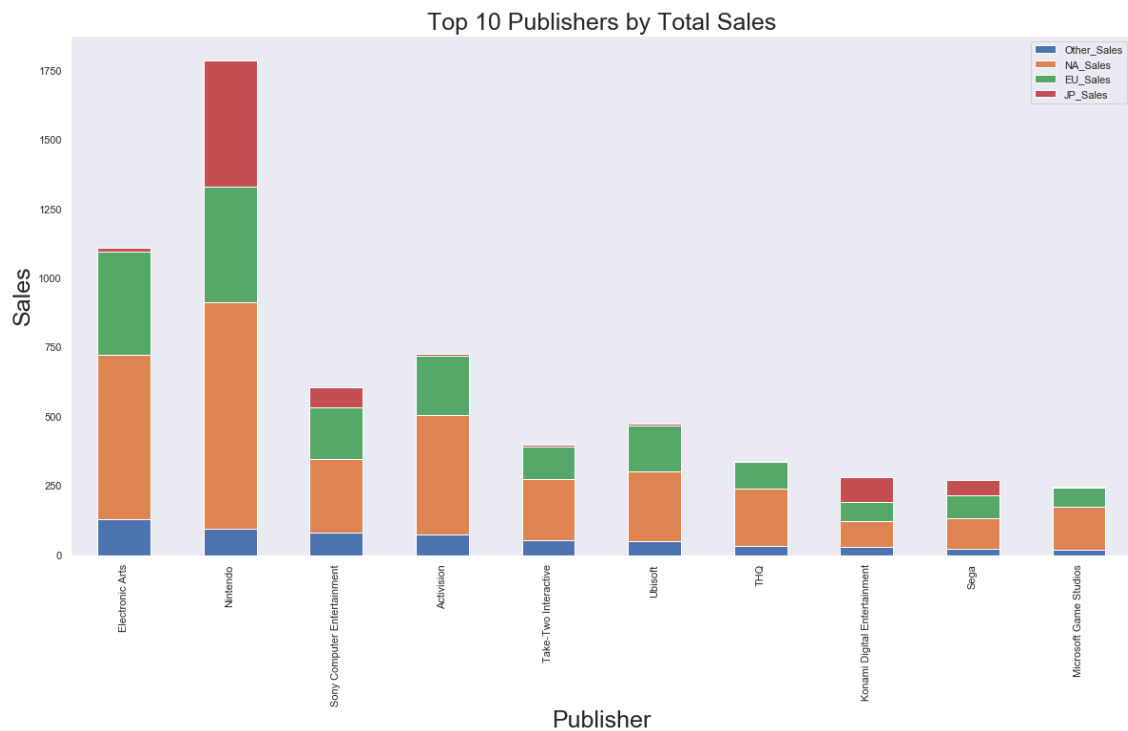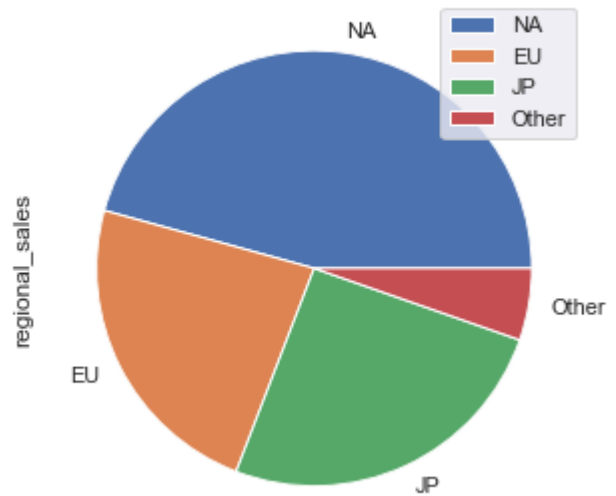

Here we clearly see that the Big 3 in gaming publishers consists of Nintendo, Electronic Arts (EA) and Sony Computer Entertainment,
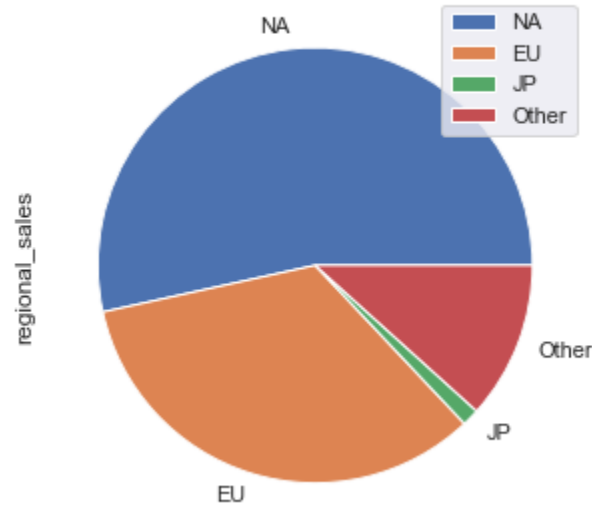Diving deeper into this Big 3 we see:
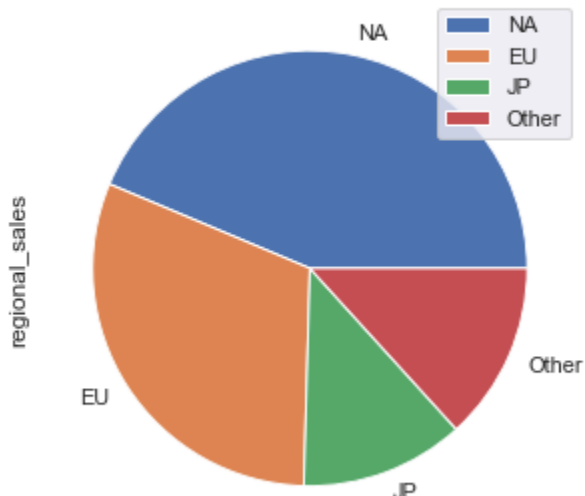
EA's Sales by region

From these charts, it's no surprise that North America is where most of the revenue is generated, owing that North America constitutes almost 38% of the Global Sales.

An interesting observation would be that, unlike in EA's regional sales, the sales figures in Japan for Sony and Nintendo are quite appreciable.

This would be due to the fact that both Sony and Nintendo are based in Japan. However, if we look at the global sales per year for each of the Big 3;

<u>Key:</u> Green: Sony
Brown: Nintendo
Blue: EA
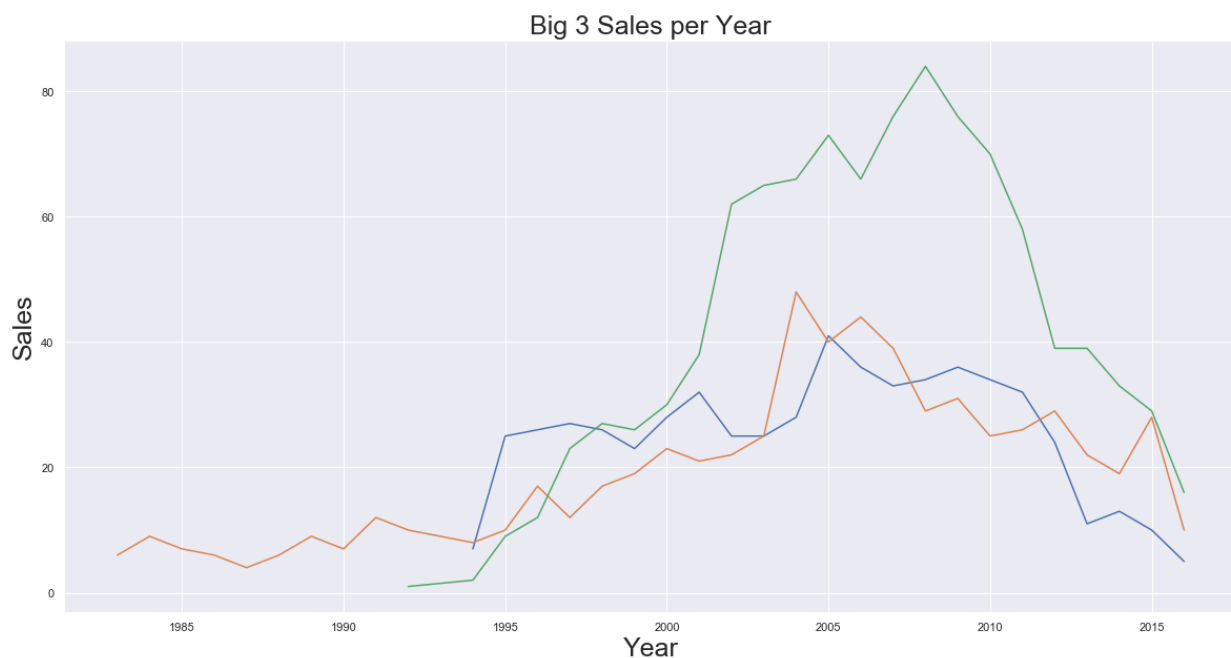
We see that Sony Entertainment skyrocketed in sales in the early 2000's, but hasn't been able to maintain that high level of sales since. Nintendo is the oldest of the three and has its revenue steadily

increasing every year. Finally EA is the 'youngest' but has enjoyed major success in the mid 2000's.

Now if we want to predict how the North American sales would improv over the years, we have the following linear regression plot.

North American Sales vs Global Sales
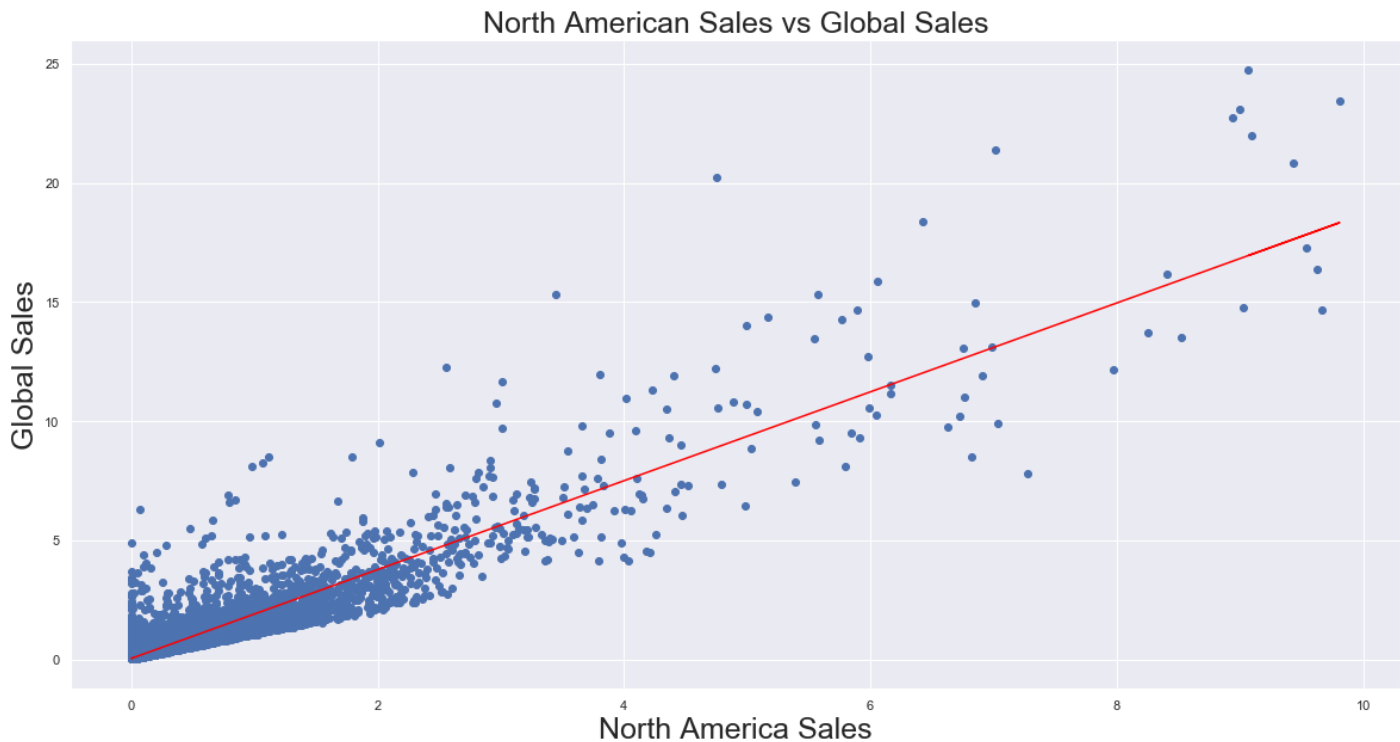
To make the plot more accurate, we used the sales data of those Global Sales between 0-10 million.

CONCLUSION

The conclusions we can draw from our dataset, after visualising all the data are:
1. The video game industry got its first real uptick in 1995.
2. The North American Region is the most popular region and hence most profitable for publishers.
3. The most popular genres are Action, Sports and Miscellaneous.
4. The most games were sold on the DS, however the most sold console was the PS2.
5. NFS: Most Wanted was sold over the most platforms, but Wii Sports is the most sold video game.
6. For a company to attain most revenue (ideally) it must develop an Action game in North America for the PC or PS4 (only other production console in the list).