

UE18CS203 – EXPLORATORY ANALYSIS ON VIDEO GAME SALES

Avinash Ratnam

Harichandana Magapu

INTRODUCTION

- The Dataset provided contained details on sales figures for all video games sold between 1980-2019.
- Using this dataset we could visualise how sales changed from year to year and how genres changed in popularity etc.
- Other conclusions based on publisher, genre trends, regional sales, publisher sales per region etc were all made.

Questions Asked

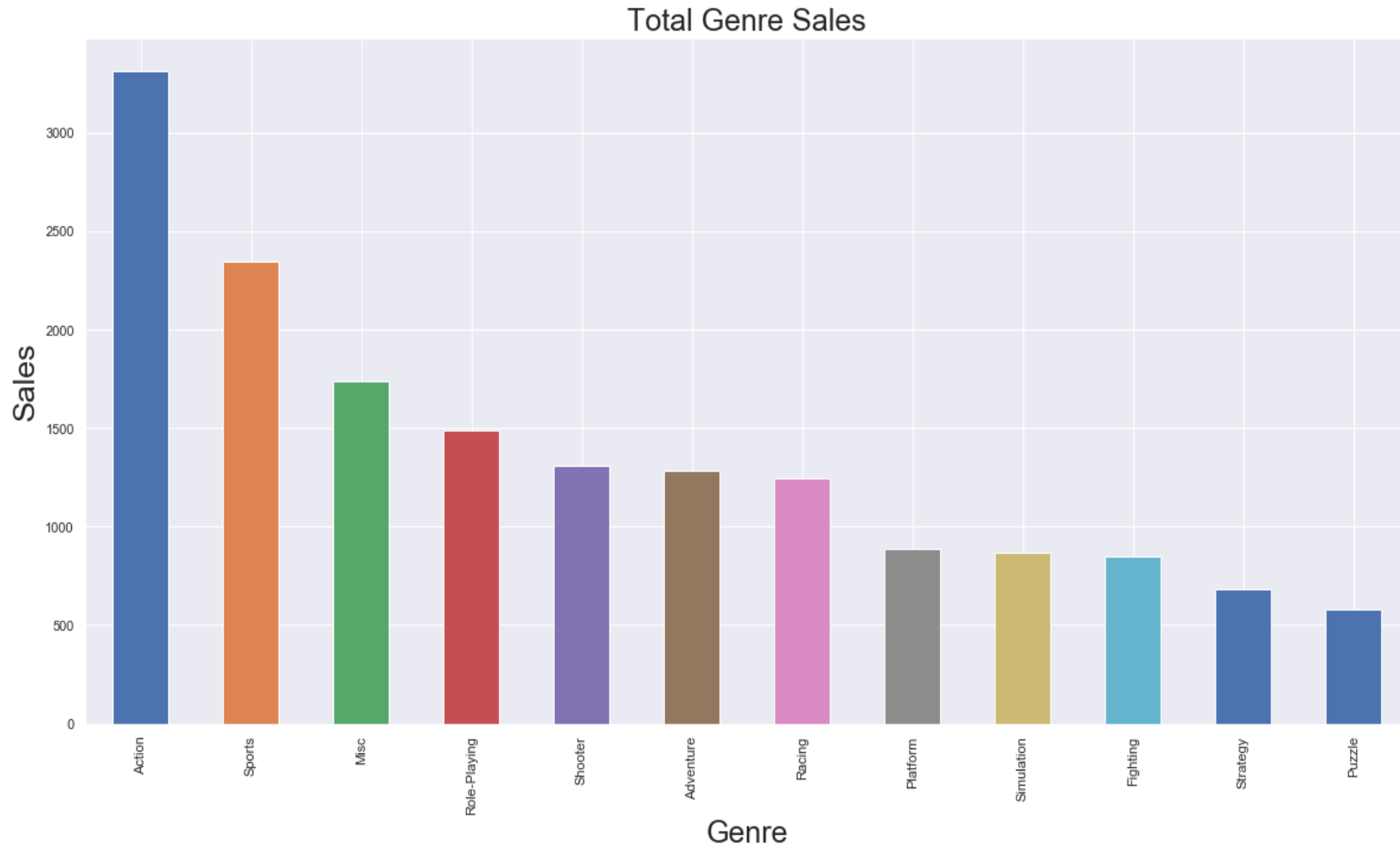
Some of the questions we thought to ourselves while creating the visualisations were:

- How did Global Sales change per year ?
- What is the most popular genre ?
- Which region prefers which genre ?
- Which is the most popular Publisher ? What are their most profitable regions ?
- Can we assume the two biggest video games have similar mean global sales ?
- Which region constitutes majority of the Global Sales ?
- How did the Big 3 publishers perform over the years ?
- Does total console sales affect the number of games sold for it ?
- Can we predict the future of Regional Sales with respect to Global Sales share ?

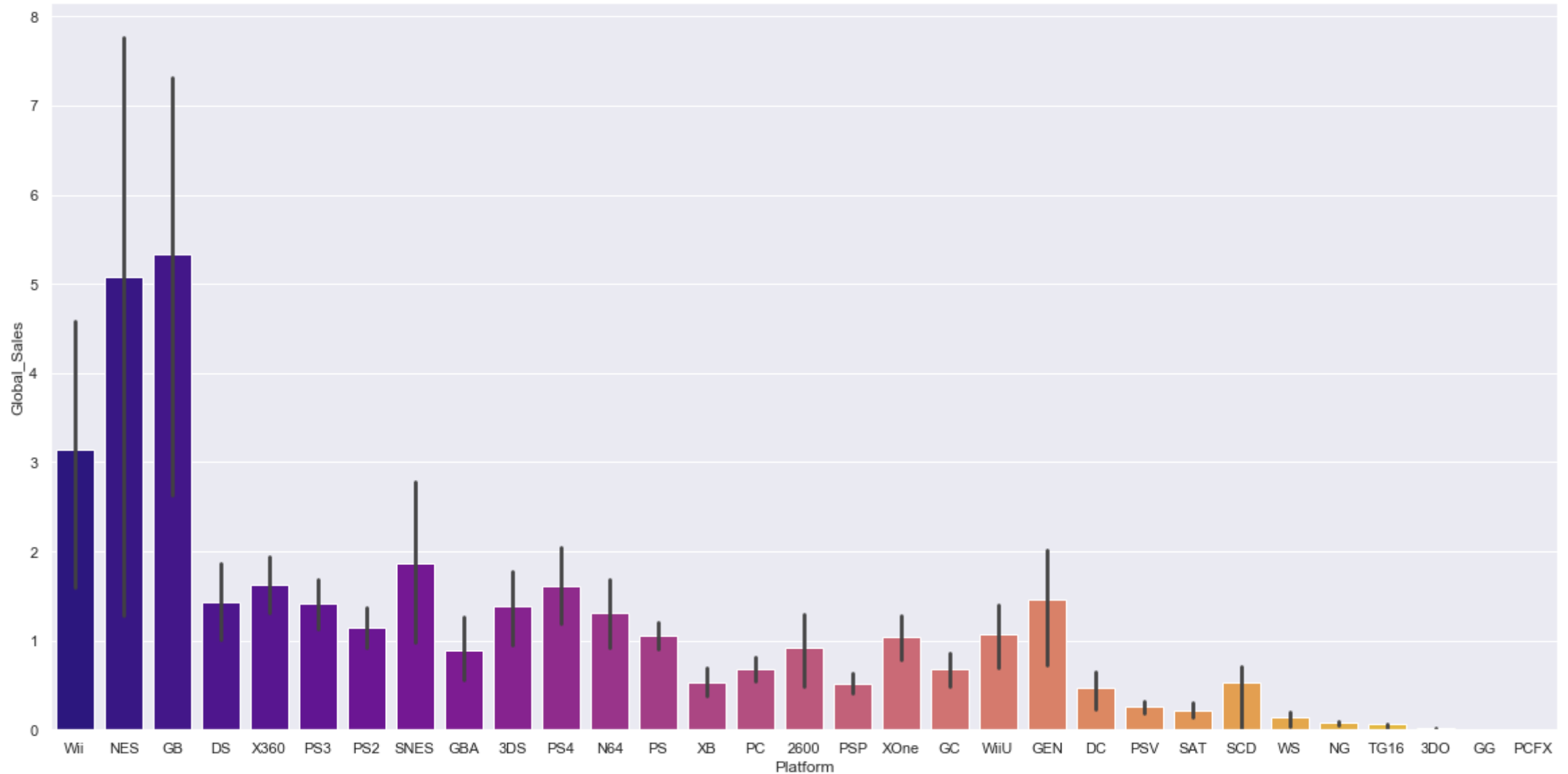
Exploratory Analysis

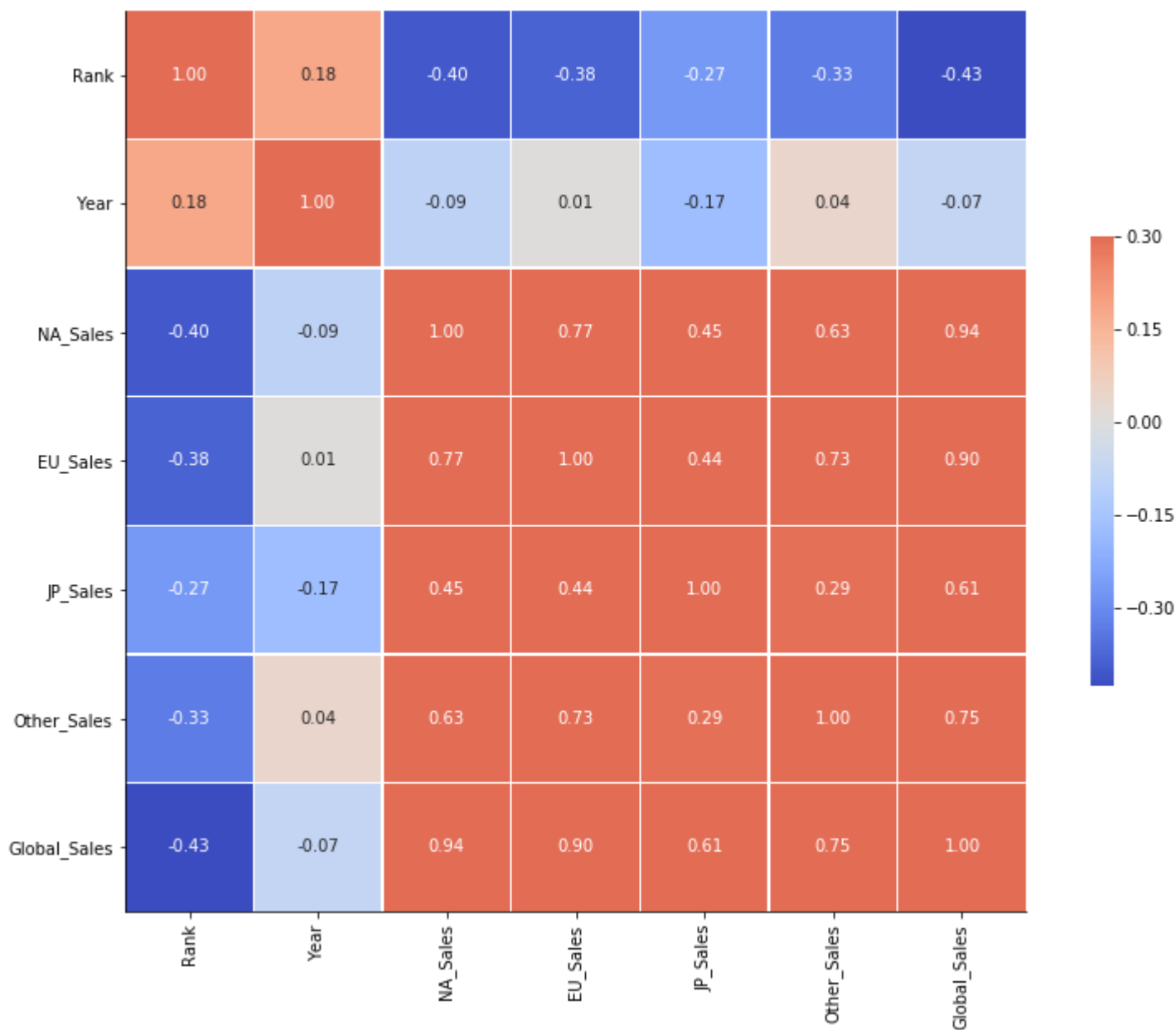


Exploratory Analysis



Exploratory Analysis





Reference

Global Sales vs NA Sales = 0.94

Global Sales vs EU Sales = 0.90

Global Sales vs JP Sales = 0.61

Exploratory Analysis

Games with the Highest number of platform releases

Name	Platform_count
Need for Speed: Most Wanted	12
FIFA 14	9
Ratatouille	9
LEGO Marvel Super Heroes	9
Angry Birds Star Wars	8
Madden NFL 07	8
Terraria	8
Monopoly	8

Publishers with most number of release titles

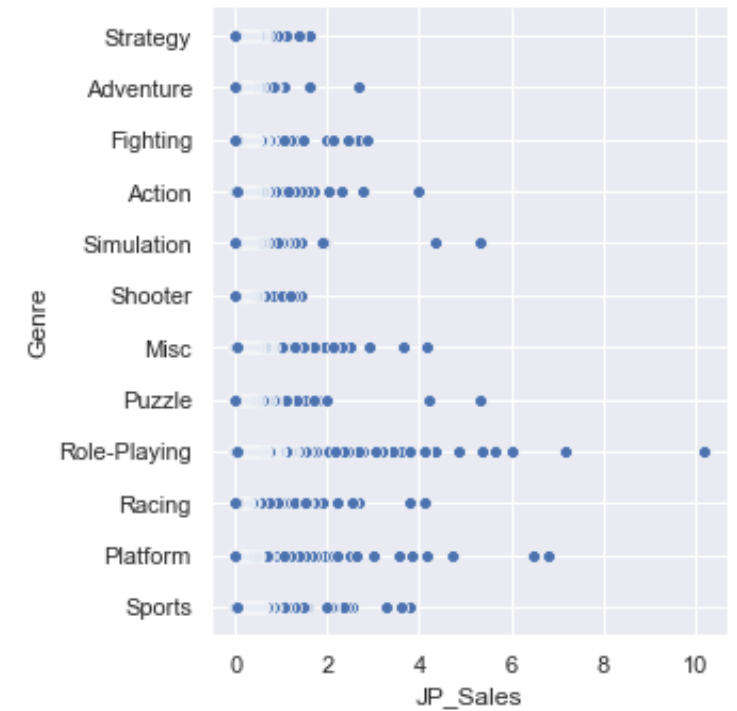
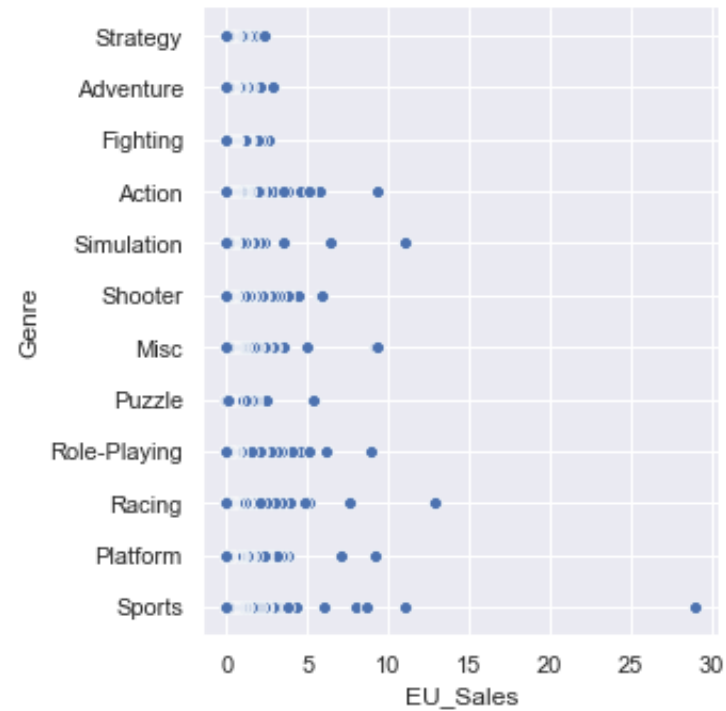
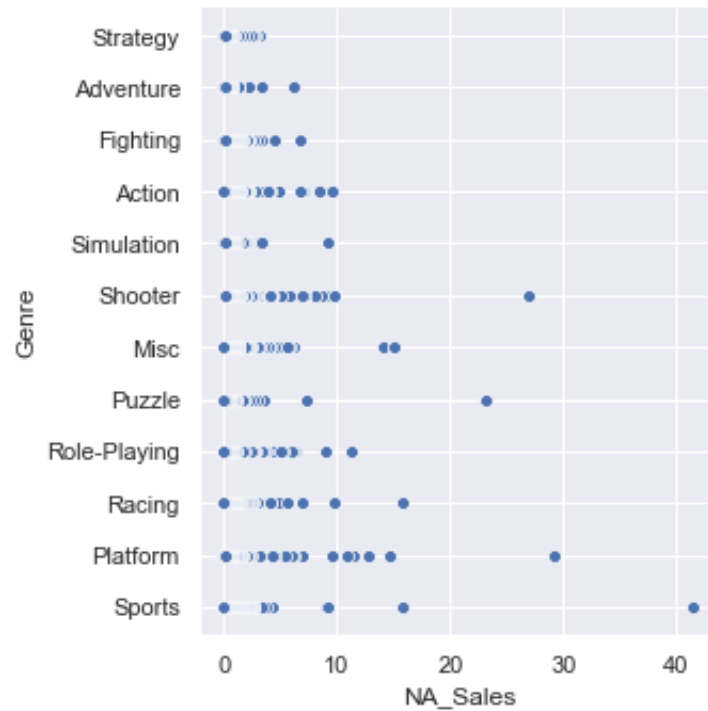
Electronic Arts	1339
Activision	966
Namco Bandai Games	928
Ubisoft	918
Konami Digital Entertainment	823
THQ	712
Nintendo	696
Sony Computer Entertainment	682
Sega	632
Take-Two Interactive	412

Big 3 Sales per Year



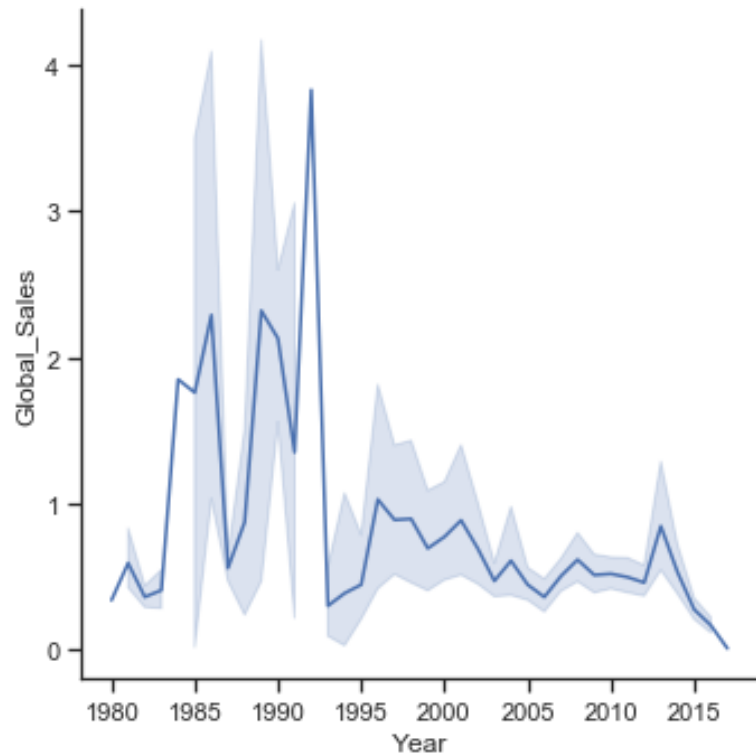
Exploratory Analysis

Genre Sales per Region (North America, Europe and Japan)

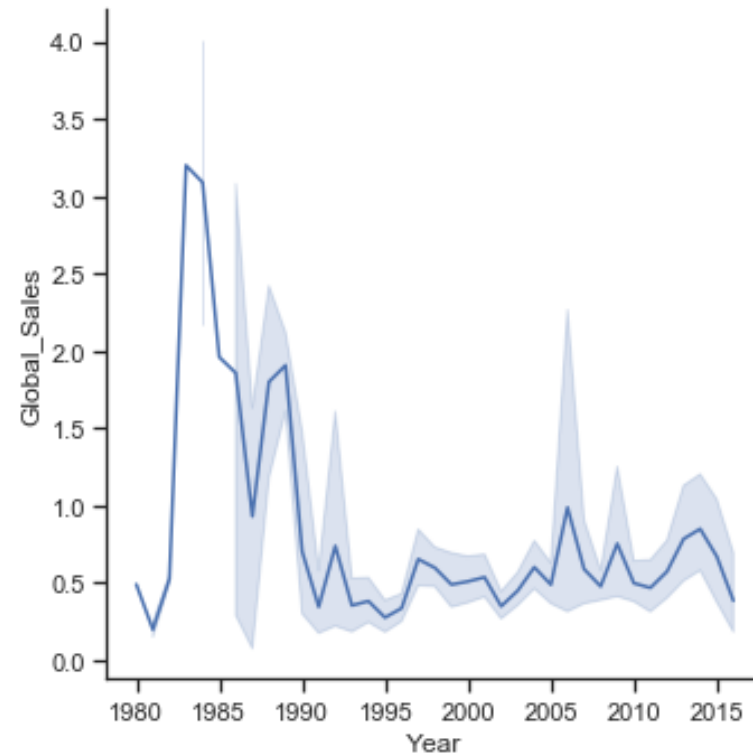


Exploratory Analysis

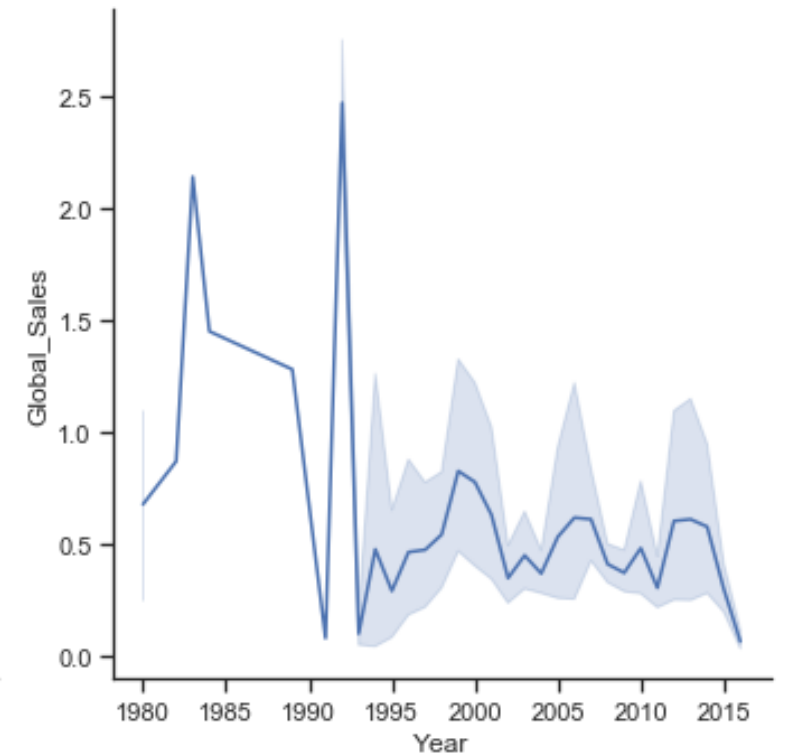
Global Sales vs Top 3 Genres



Action Games

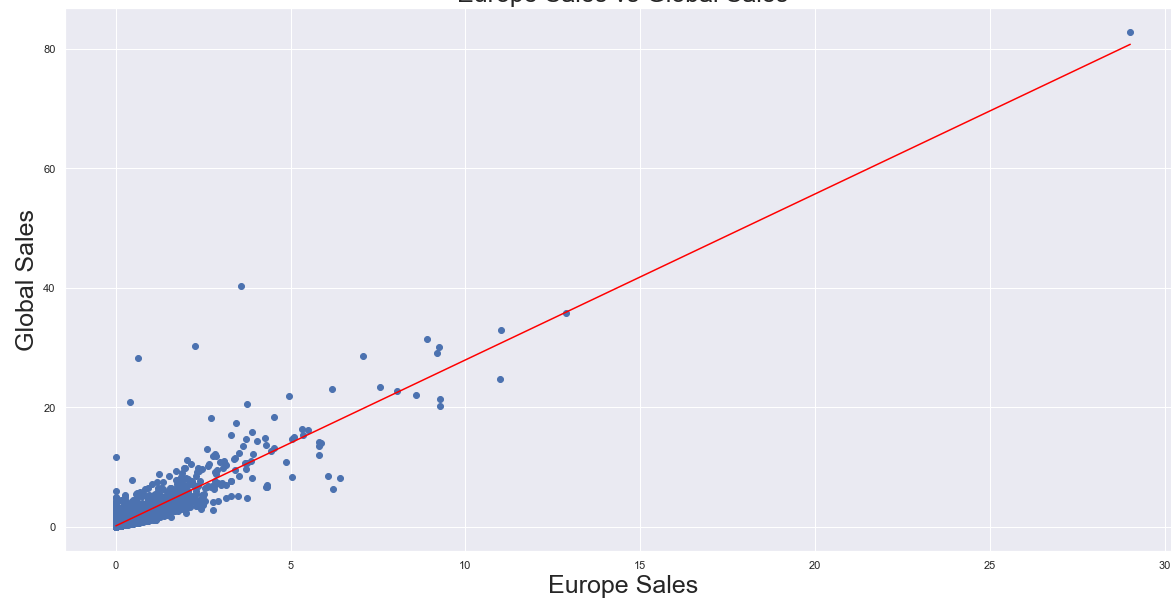


Sports Games

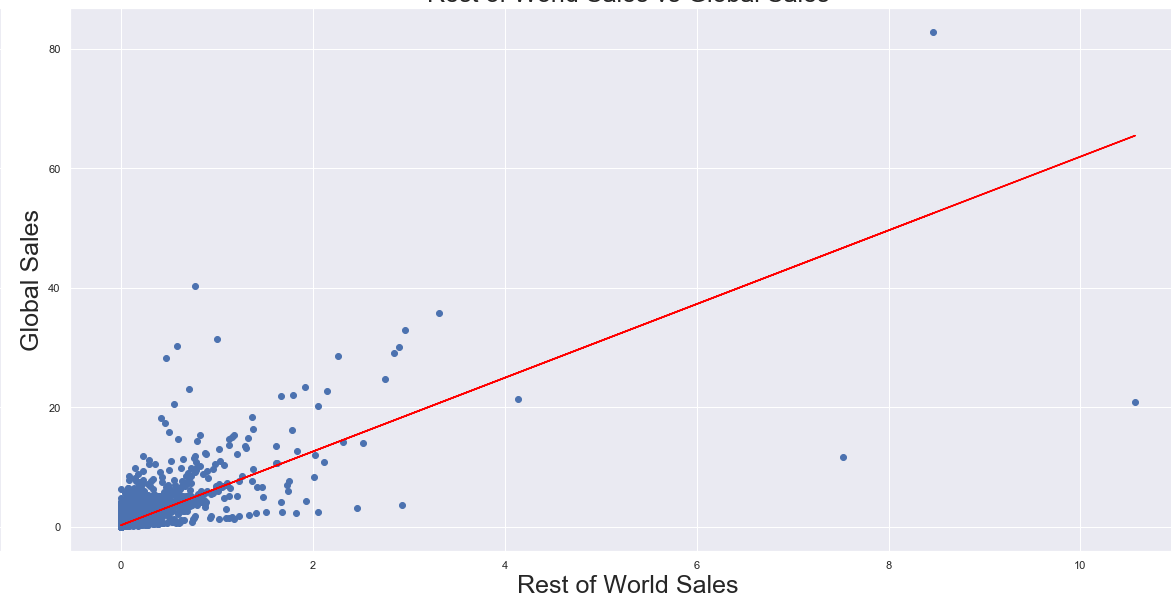


Miscellaneous Games

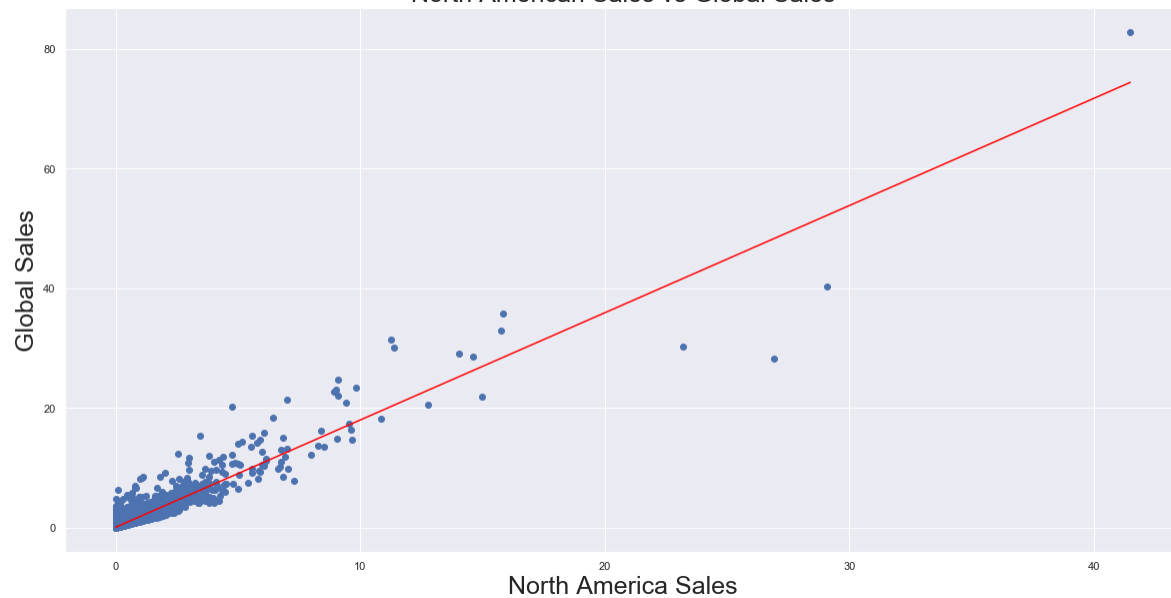
Europe Sales vs Global Sales



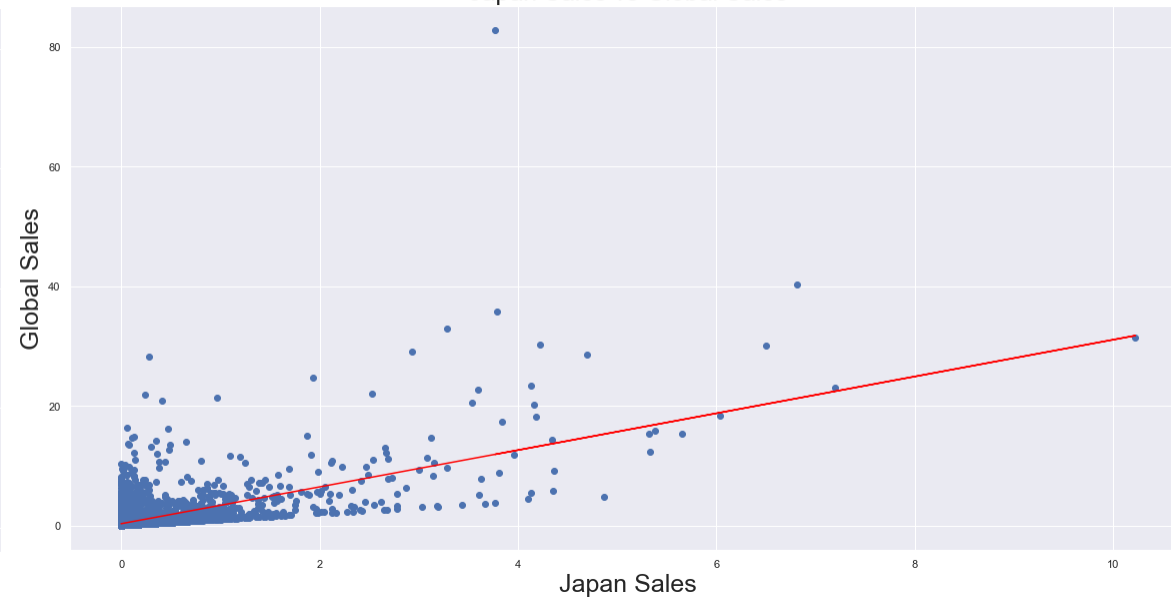
Rest of World Sales vs Global Sales



North American Sales vs Global Sales



Japan Sales vs Global Sales



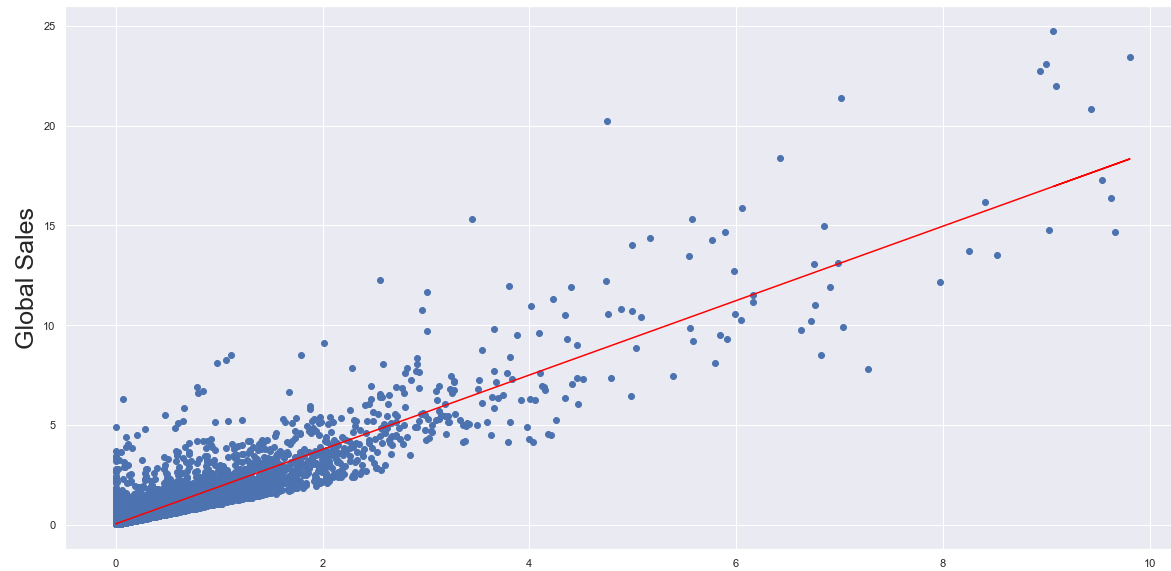
Exploratory Analysis

From the scatter plots in the previous slide, we see that even though the regression values are very high (around 0.94 between NA Sales and Global Sales) the linear regression model is not a good fit.

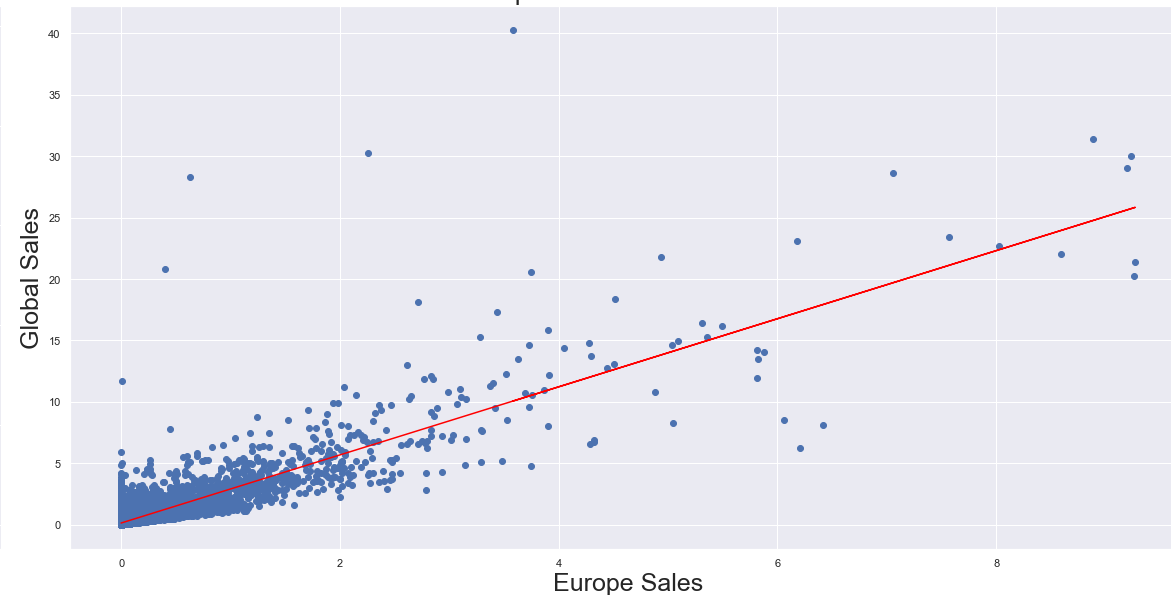
This is because, the standard error is not constant (most of the data points are between 0-1)

However, if we consider only Regional Sales between 0-10 million, we see that linear regression is indeed a good fit for North America and Europe Sales.

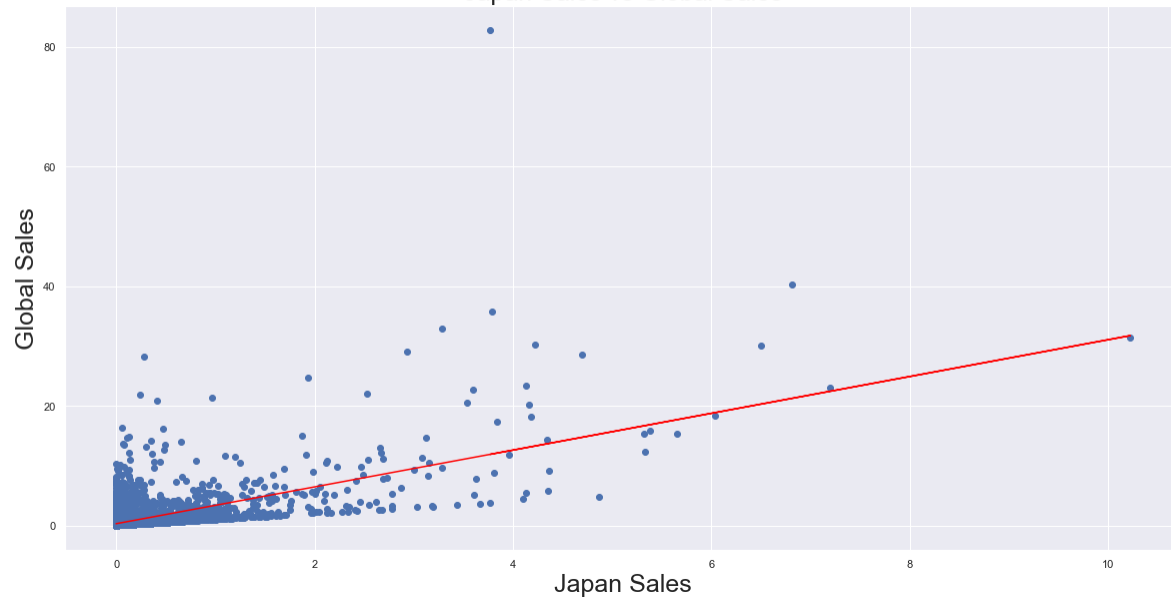
North American Sales vs Global Sales



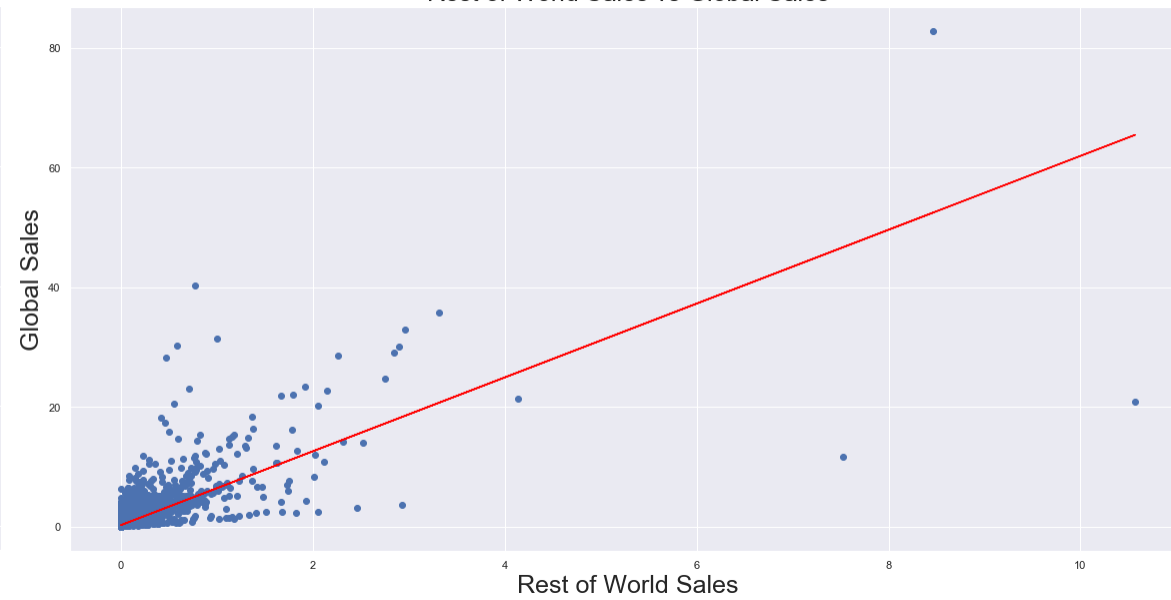
Europe Sales vs Global Sales



Japan Sales vs Global Sales



Rest of World Sales vs Global Sales



Hypothesis Testing

```
pokemon=nintendo.loc[nintendo['Name'].str.contains('Pokemon')]
pokemon.append(nintendo.loc[nintendo['Name'].str.contains('Pokémon')])
pokemon.sort_values(by='Rank')
```

```
mario=nintendo.loc[nintendo['Name'].str.contains('Mario')]
pokemon.append(nintendo.loc[nintendo['Name'].str.contains('Super Smash Bros')])
mario.sort_values(by='Rank')
```

```
p_sample=pokemon.sample(frac =.25)
m_sample=mario.sample(frac =.25)
from statsmodels.stats import weightstats as stests
#H0: Pokemon and Mario games have equal mean global sales
#H1: Pokemon and Mario games don't have equal mean global sales
ztest ,pval = stests.ztest(p_sample['Global_Sales'], x2=m_sample['Global_Sales'], value=0,alternative='two-sided')
print('P-value: ',float(pval))
if pval<0.05:
    print("reject null hypothesis")
else:
    print("fail to reject null hypothesis")
```

P-value: 0.9135569100295993
fail to reject null hypothesis

Conclusions

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).