Project: Video Game Sales Analysis
May 2023
Project 1 Group 1

Project Members:
Raj Agrawal
Van Eisele
John Banowsky

Data: Sales of Video Games
Analyze sales data from more than 11,000 games from 1980 to 2020.
https://www.kaggle.com/datasets/arslanali4343/sales-of-video-games

Project Goal: Identify what makes profitable games.

Project Questions:
Are there any regions that help dictate successful game sales?
Does genre have any impact on sales?
Do certain publishers do better in video game sales?
Are there any differences in gaming platforms sales?

Analysis Summary:
Data cleaning - cleaned headers, included Platform Company and Decade Columns,
drop null years and post 2015 years.

Used categorical graphs to identify leading platforms and genres by overall, decade and
platform company.
Used correlations and regressions to analyze the relationship between North American
sales vs other regions.
Calculated outliers by genre for 2010s.

Analysis Results:
Sales of games bundled with consoles or hardware outperform single game sales.
Long running franchises and series have a built up following and maintain repeat buyers
with sustained high value sales.
New unique games with special novelties also are a good way to break into the market.
Genre comparisons show that shooters have a higher probability of success compared
to other genres in North America.
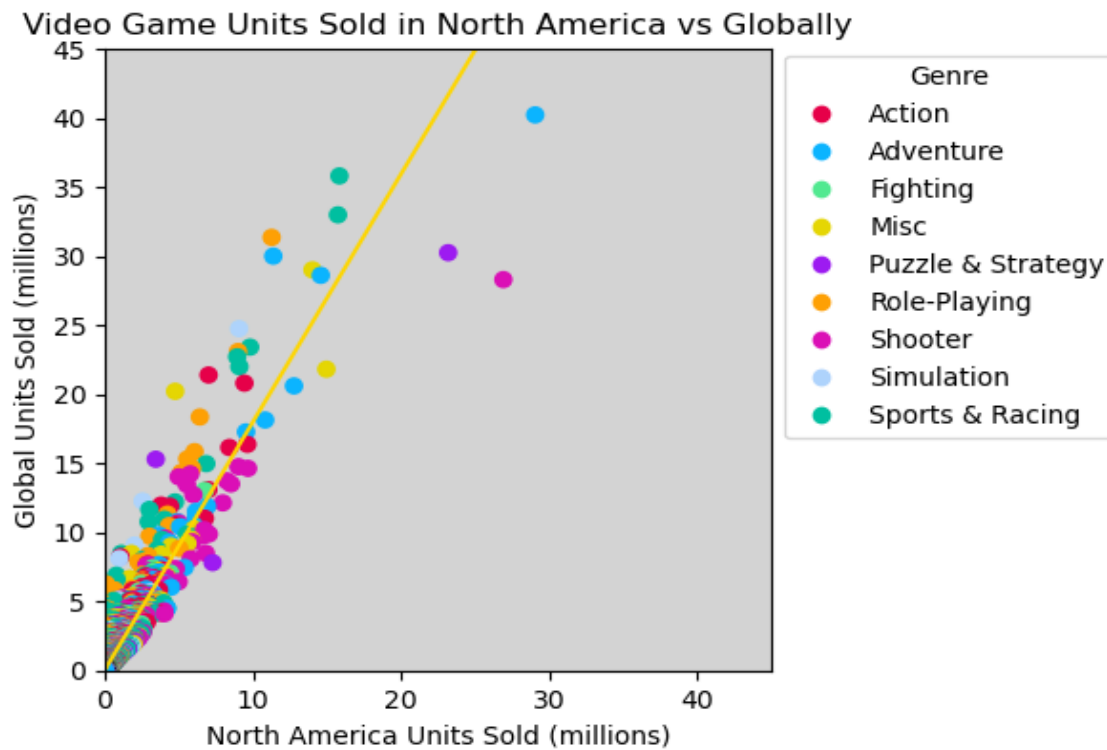Role-Playing games are successful in most regions.

THE PROCESS:

Game Sales Trends:

First, we ran a basic line graph modeling the game sales per unit with the 4 different regions to get a grasp of the information.

Video Game Units Sold by Year
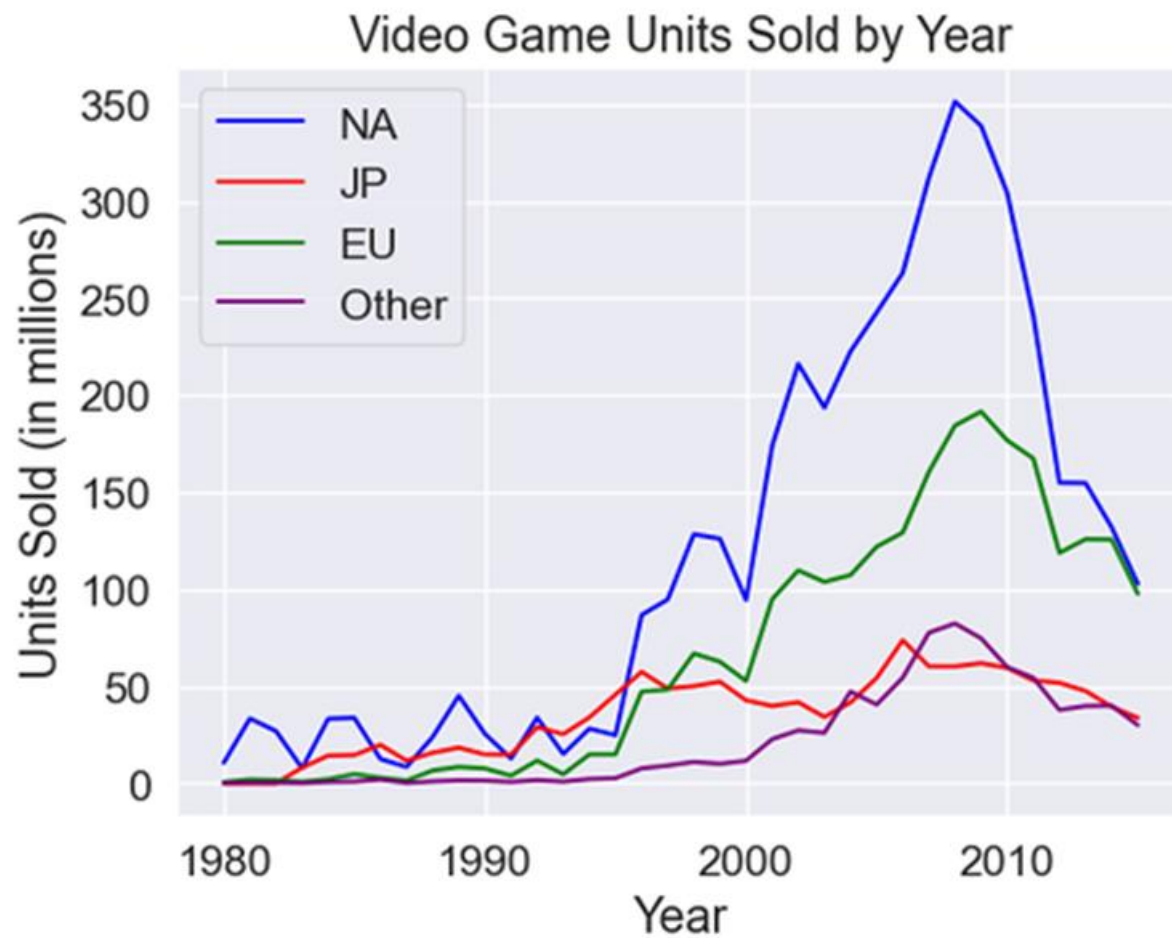
We noticed that the North American sales was clearly leading the pack in the most recent years, so we decided to investigate any correlations between North American Sales and Global sales separating out genres for a closer look.

Video Game Units Sold in North America vs Globally

Genre
- Action
- Adventure
- Fighting
- Misc
- Puzzle & Strategy
- Role-Playing
- Shooter
- Simulation
- Sports & Racing

We then ran a correlation model to get confirmation on a correlation between North American Sales and Global Sales.

```
                          OLS Regression Results
==============================================================================
Dep. Variable:          Global_Sales   R-squared:                       0.886
Model:                            OLS   Adj. R-squared:                  0.886
Method:                 Least Squares   F-statistic:                 1.247e+05
Date:                Wed, 17 May 2023   Prob (F-statistic):               0.00
Time:                        19:05:34   Log-Likelihood:                 -12606.
No. Observations:               15979   AIC:                         2.522e+04
Df Residuals:                   15977   BIC:                         2.523e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0637      0.004     14.382      0.000       0.055       0.072
NA_Sales       1.7935      0.005    353.115      0.000       1.784       1.803
==============================================================================
Omnibus:                     9342.128   Durbin-Watson:                   1.927
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         37165632.327
Skew:                           1.074   Prob(JB):                         0.00
Kurtosis:                     239.256   Cond. No.                         1.42
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Further Exploration into Sales Trends:

Now that we established that there is a correlation between North American sales and Global sales, we took a deeper dive to look at any other correlations between North America and the other regions and breaking down the genres to get a better view of the marketplace.

Video Game Units Sold in North America vs Europe

Video Game Units Sold in North America vs Japan
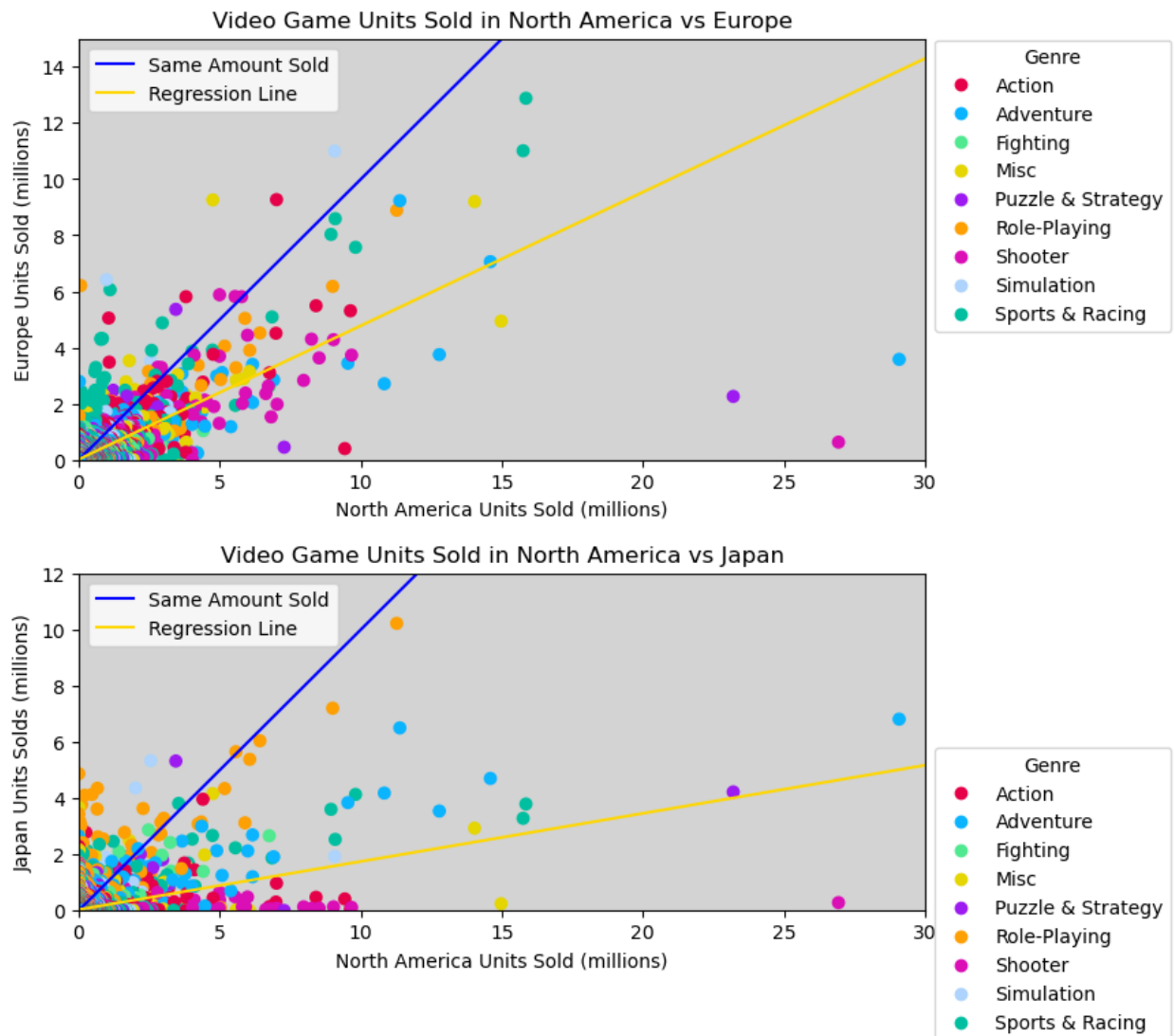
	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
Year	1.000000	-0.085347	0.011394	-0.170121	0.047488	-0.069223
NA_Sales	-0.085347	1.000000	0.770262	0.451618	0.634225	0.941499
EU_Sales	0.011394	0.770262	1.000000	0.437859	0.725572	0.903753
JP_Sales	-0.170121	0.451618	0.437859	1.000000	0.290955	0.613358
Other_Sales	0.047488	0.634225	0.725572	0.290955	1.000000	0.747362
Global_Sales	-0.069223	0.941499	0.903753	0.613358	0.747362	1.000000
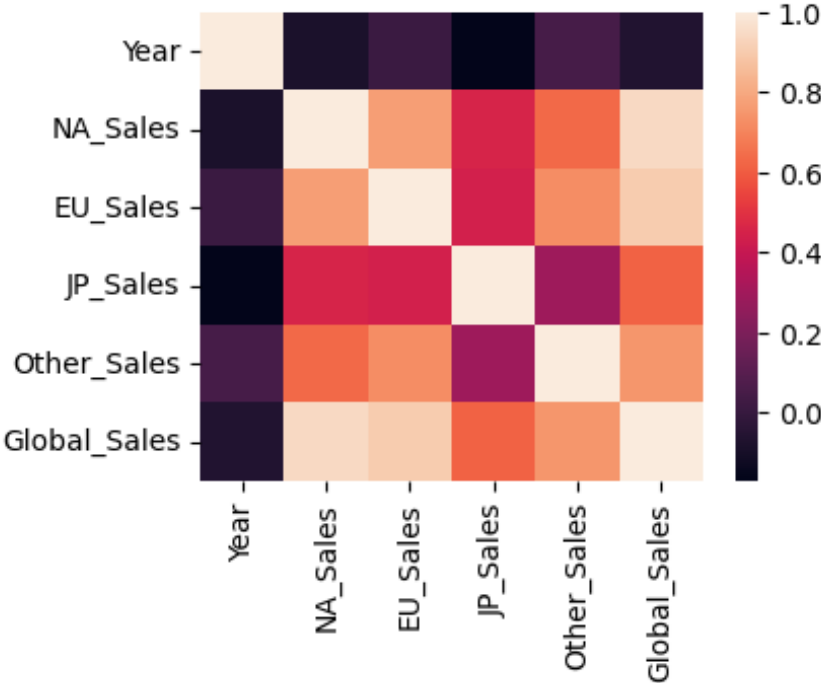
Sales by Genre:

Now that we have established correlation between regions, we wanted to look at genre sales over the years and we narrowed down the genres we wanted to focus on for our game production. We did this by trying to plot some box and violin charts and calculating quartiles. We wanted to focus on Shooters, RPGs and Sports and Racing Games.

Outlier Exploration for 2010s

There were 1320 total Action games with 143 considered outliers. 10.8% were outliers selling greater than 0.5 million units.

There were 650 total Adventure games with 91 considered outliers. 14.0% were outliers selling greater than 0.18 million units.

There were 183 total Fighting games with 12 considered outliers. 6.6% were outliers selling greater than 0.58 units.

There were 545 total Misc games with 51 considered outliers. 9.4% were outliers selling greater than 0.45 units.

There were 274 total Puzzle & Strategy games with 23 considered outliers. 8.4% were outliers selling greater than 0.2 units.

There were 516 total Role-Playing games with 64 considered outliers. 12.4% were outliers selling greater than 0.4 million units.
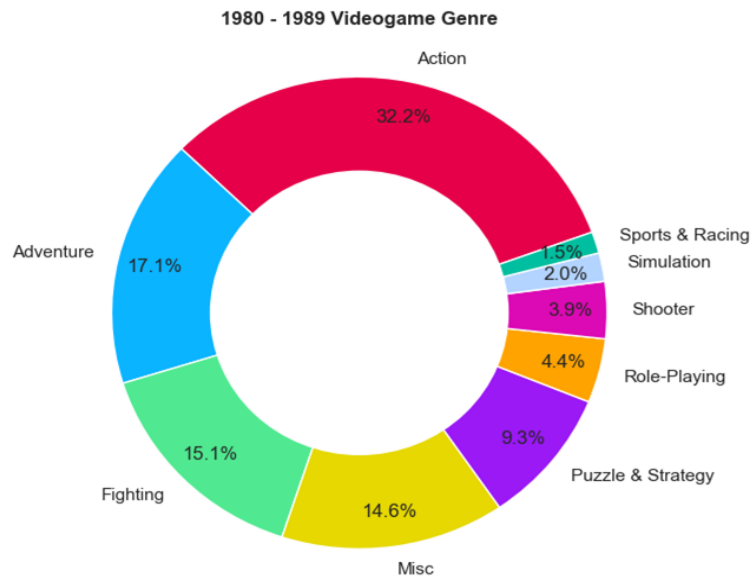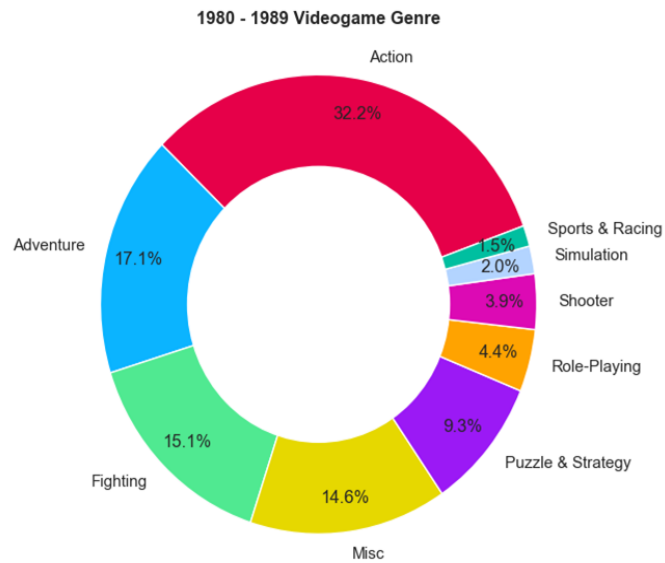
There were 363 total Shooter games with 45 considered outliers. 12.4% were outliers selling greater than 1.24 million units.

There were 200 total Simulation games with 24 considered outliers. 12.0% were outliers selling greater than 0.31 units.

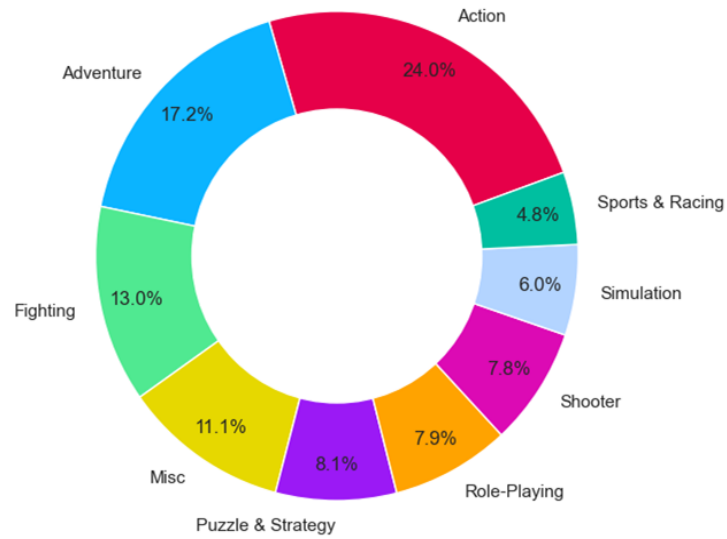**There were 746 total Sports & Racing games with 78 considered outliers. 10.5% were outliers selling greater than 0.67 million units.**
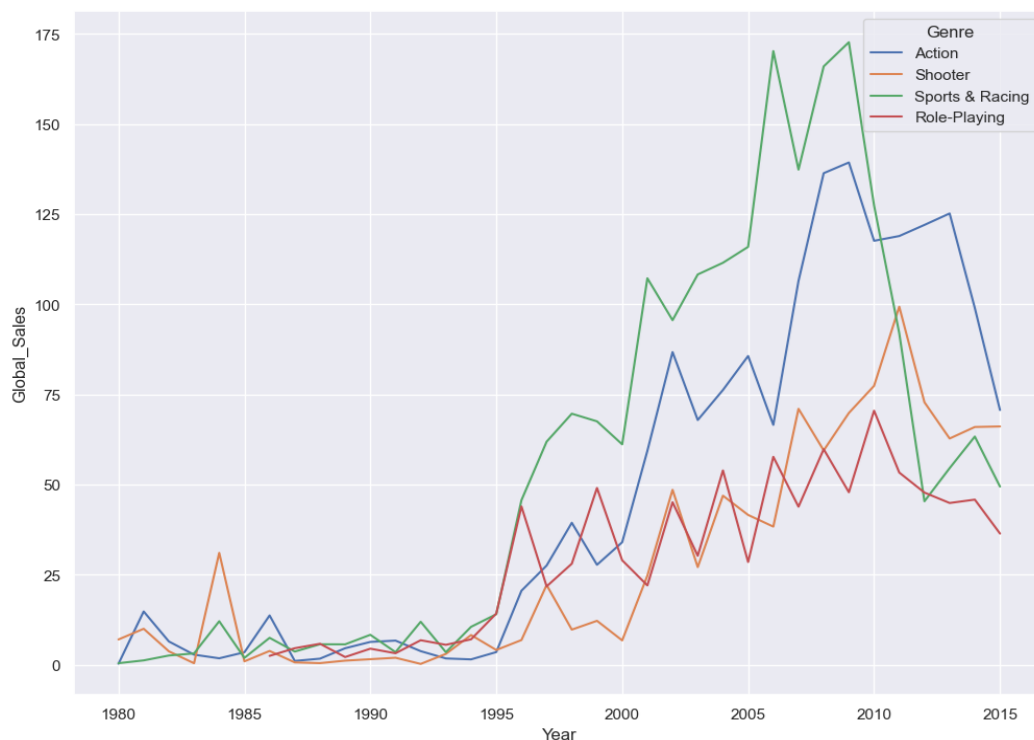
We looked at genre sales over the decades for more domain expertise and to see if there were any trends. Such as big growth in certain genres and if there might be a reason, such as just basic hardware improvements.

1980 - 1989 Videogame Genre

1980 - 1989 Videogame Genre

2000 - 2009 Videogame Genre

Action 24.0%
Sports & Racing 4.8%
Simulation 6.0%
Shooter 7.8%
Role-Playing 7.9%
Puzzle & Strategy 8.1%
Misc 11.1%
Fighting 13.0%
Adventure 17.2%

2010 - 2015 Videogame Genre

Action 27.5%
Sports & Racing 3.8%
Simulation 4.2%
Shooter 5.7%
Role-Playing 7.6%
Puzzle & Strategy 10.8%
Misc 11.4%
Fighting 13.6%
Adventure 15.6%
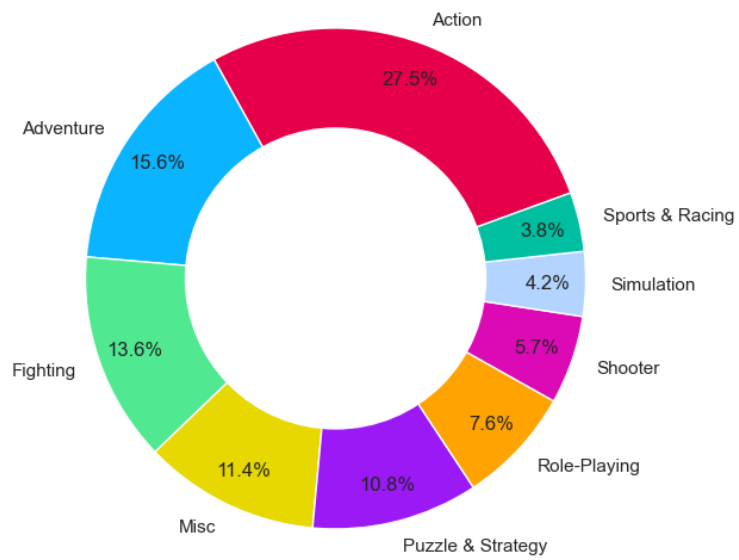
Platform Company Exploration:

After narrowing down our target regions and which genre to make, we wanted to look to see which platform company to go with in case we were not making a cross platform game or if making a cross platform game are there any we should target over others.
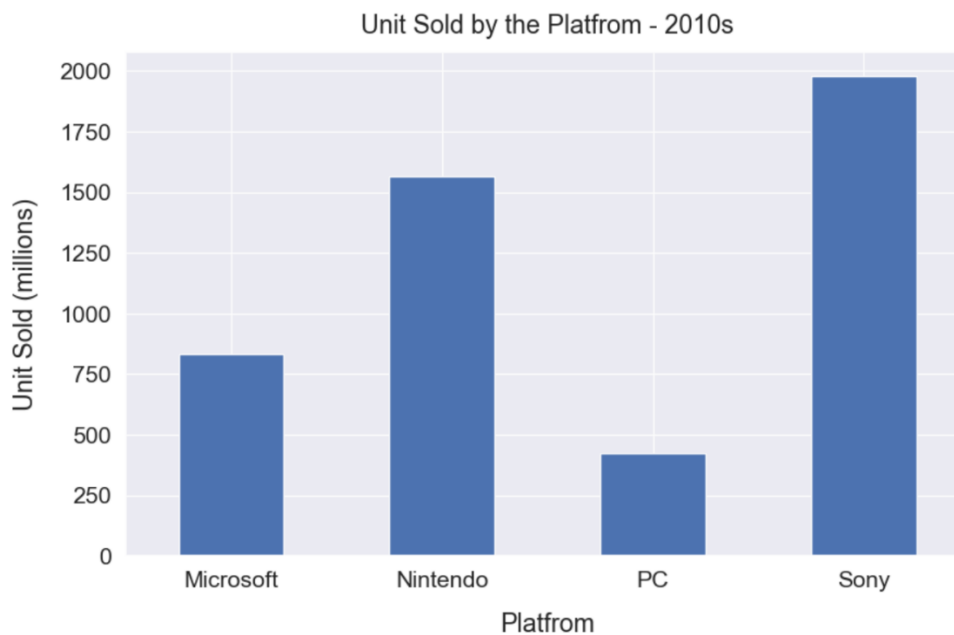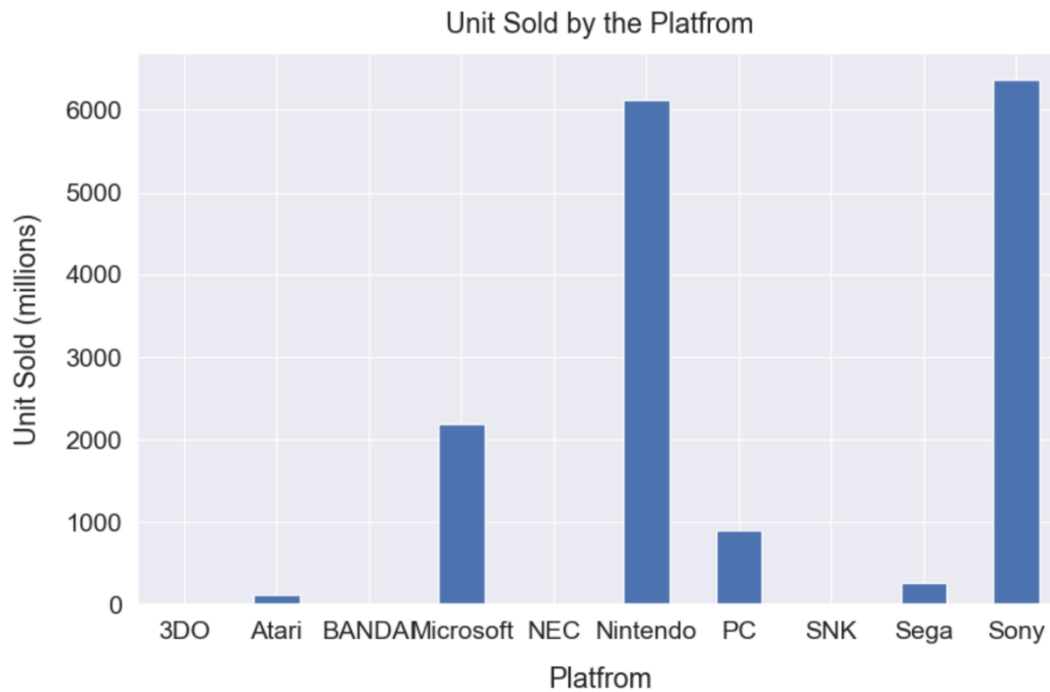
It became very apparent that the leading platform companies are Sony, Nintendo, Microsoft, and PC and in that order. With this information we now wanted to do a breakdown of each platform and the genres that are sold on those platforms.

2010s Genres on Microsoft

Action 30.3%
Adventure 7.2%
Fighting 5.4%
Misc 10.6%
Puzzle & Strategy 0.6%
Role-Playing 5.2%
Shooter 15.7%
Simulation 2.6%
Sports & Racing 22.4%

2010s Genres on PC

Action 23.7%
Adventure 8.7%
Fighting 0.6%
Misc 2.1%
Puzzle & Strategy 15.0%
Role-Playing 10.8%
Shooter 15.9%
Simulation 10.5%
Sports & Racing 12.6%

2010s Genres on Sony

Genre	Percentage
Action	31.7%
Sports & Racing	13.7%
Simulation	1.5%
Shooter	6.6%
Role-Playing	15.1%
Puzzle & Strategy	2.6%
Misc	6.2%
Fighting	5.4%
Adventure	17.3%

2010s Genres on Nintendo

Genre	Percentage
Action	32.5%
Sports & Racing	10.2%
Simulation	4.4%
Shooter	2.4%
Role-Playing	10.8%
Puzzle & Strategy	8.2%
Misc	17.4%
Fighting	2.2%
Adventure	11.9%
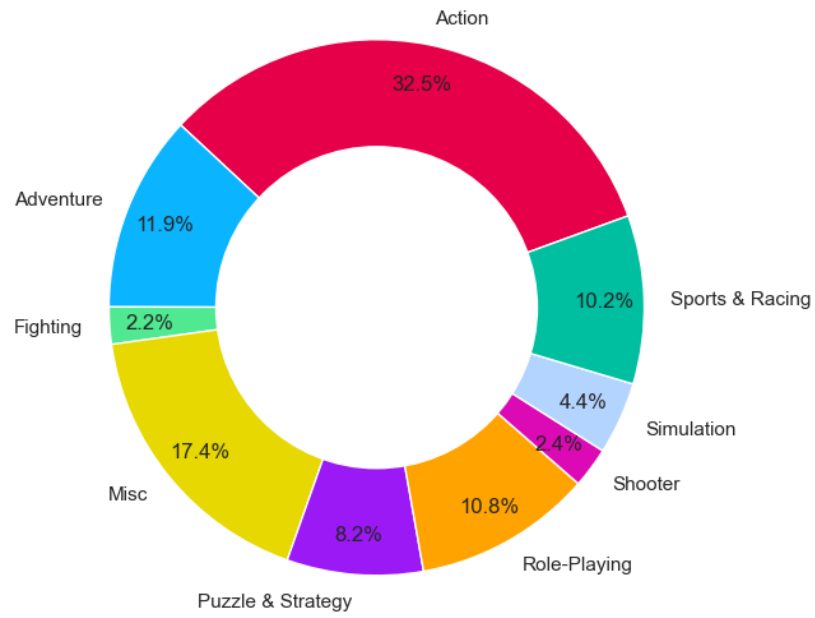
After exploring the platform companies, we now went back to look at correlations between the genre sales and regions in the genres we were interested in. We also wanted to identify outlier games to get a better understanding of what type or games are popular.

Role-Playing Units Sold in North America vs Japan
Linear Regression: y = 0.61 x + 0.10   R-squared: 0.402

Role-Playing Units Sold in North America vs Europe
Linear Regression: y = 0.64 x – 0.01   R-squared: 0.767

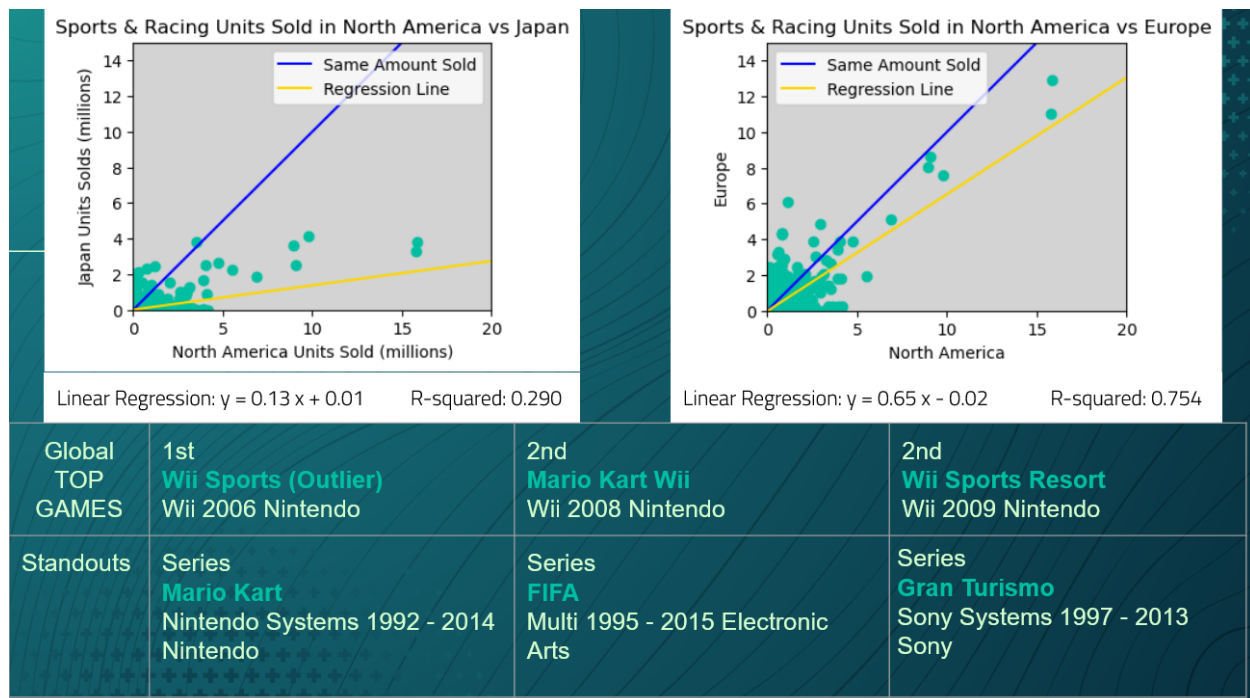| Global TOP GAMES	1st Pokemon Red / Blue (Green) Game Boy 1996 Nintendo	2nd Pokemon Gold / Silver Game Boy 1999 Nintendo	3rd Pokemon Diamond / Pearl DS 2006 Nintendo
Standouts	North America The Elder Scrolls V: Skyrim Multi 2011 Bethesda Softworks	Japan Monster Hunter Freedom 3 PSP 2010 Sony	Europe World of Warcraft PC 2004 Activision

Shooter Units Sold in North America vs Japan
Linear Regression: y = 0.02 x + 0.02   R-squared: 0.048

Shooter Units Sold in North America vs Europe
Linear Regression: y = 0.30 x + 0.10   R-squared: 0.455

Global TOP GAMES	1st Dunk Hunt (Outlier) Game Boy 1996 Nintendo	2nd Call of Duty: Modern Warfare 3 Multi 2011 Activision	2nd Call of Duty: Black Ops Multi 2010 Activision
Standouts	Japan Splatoon WiiU 2015 Nintendo	Series Call of Duty: Modern Warfare & Black Ops Multi 2007- 2015 Activision	Series Halo Xbox Systems 2001 - 2012 Microsoft Game Studios

Sports & Racing Units Sold in North America vs Japan

Linear Regression: y = 0.13 x + 0.01   R-squared: 0.290

Sports & Racing Units Sold in North America vs Europe

Linear Regression: y = 0.65 x - 0.02   R-squared: 0.754

Global TOP GAMES	1st **Wii Sports (Outlier)** Wii 2006 Nintendo	2nd **Mario Kart Wii** Wii 2008 Nintendo	2nd **Wii Sports Resort** Wii 2009 Nintendo
Standouts	Series **Mario Kart** Nintendo Systems 1992 - 2014 Nintendo	Series **FIFA** Multi 1995 - 2015 Electronic Arts	Series **Gran Turismo** Sony Systems 1997 - 2013 Sony

Next up we did some T-Test. We used t-tests to determine whether the average game sales of shooter games in North America was greater compared to other genres and regions. The p-values indicate that the average games sales of shooters are statistically significant.

T-Test Comparing North American Shooter Sales (Code)

```
1  pop_x = shooter.NA_Sales
2  pop_y = rpg.NA_Sales
3  print(pop_x.mean())
4  print(pop_x.var())
5  print(len(pop_x))
6  print(pop_y.mean())
7  print(pop_y.var())
8  print(len(pop_y))
```

```
0.45417600000000014
1.501293035452347
1250
0.22750874737578727
0.469441408361786
1429
```

```
1  pop_x = shooter.NA_Sales
2  pop_y = sports.NA_Sales
3  print(pop_x.mean())
4  print(pop_x.var())
5  print(len(pop_x))
6  print(pop_y.mean())
7  print(pop_y.var())
8  print(len(pop_y))
```

```
0.45417600000000014
1.501293035452347
1250
0.29438940092165894
0.9234260171506634
3472
```

```
1  pop_x = shooter.NA_Sales
2  pop_y = shooter.EU_Sales
3  print(pop_x.mean())
4  print(pop_x.var())
5  print(len(pop_x))
6  print(pop_y.mean())
7  print(pop_y.var())
8  print(len(pop_y))
```

```
0.45417600000000014
1.501293035452347
1250
0.2422
0.3093768214571657
1250
```

```
1  st.ttest_ind(pop_x.values, pop_y.values, equal_var=False)
```

```
Ttest_indResult(statistic=5.6097667673362075, pvalue=2.3336277662001025e-08)
```

T-Test Comparing North American Shooter Sales

	NA Shooters	vs EU Role-Playing	vs JP Role-Playing	vs EU Shooters	vs JP Shooters	vs EU Sports & Racing	vs JP Sports & Racing
Mean (million)	0.454	0.130	0.243	0.242	0.030	0.173	0.055
Variance	1.50	0.25	0.43	0.31	0.01	0.52	0.06
Total	1250	1429	1429	1250	1250	3472	3472
P-Value	Null	≈0 (6.08e-18)	≈0 (5.58e-8)	≈0 (2.95e-8)	≈0 (2.03e-32)	≈0 (3.24e-14)	≈0 (6.26e-29)

	Mean (million)	Variance	Total	P-Value
NA Shooters	0.454	1.50	1250	Null
vs NA Role-Playing	0.294	0.92	3472	0.0000316
vs NA Sports & Racing	0.227	0.47	1429	0.00000000794

Predictions:

A game needs to be successful in North America to be successful globally.

The novelty of an RPG/Shooter that is on multiple platforms has a high chance of success.

Upcoming game that matches our criteria: Palworld https://www.pocketpair.jp/palworld

Future Study:

Further explore multi-platform games

Correlation of Handhelds and RPGS

Influence of modern gaming post 2015

Combine data from long running games series [FIFA, Call of Duty, Halo, Madden, Pokémon, Mobile games]

Data Bias:

Selection bias (data doesn't include all games)

Availability Bias (Date after 2015 is not available)

Streaming and Downloads in recent times have substituted physical units.

Emergence of new video game technology: Mobile Games in late 2010s