Video Game Sales Analysis

John Ruddick

MDA 620: Data-driven Decision-making

Professor Singh

11 December 2022

Overview/Background:

This project will analyze the Video Game sales across all platforms from 1980 through 2016, as well as predict future Sales trends and patterns using a public dataset obtained from Kaggle. Each row in the data set represents an individual video game released during the time frame. The data is categorized in columns by year of release, genre, publisher, and platform. Sales are broken out by region (North America, Japan, Europe).

Column	Description	Expected Values
Name	Individual Video Game titles	Titles
Platform	Console Game is played on	Platform Titles
Year of Release	Year game was released	Date(Year)
Genre	Genre Category Game falls into	Genre Title
Publisher	Publisher of video game	Publisher TRitle
NA Sales	North American Sales	Sales Value
Eu Sales	Europe Sales	Sales Value
Jp Sales	Japan Sales	Sales Value
Other Sales	Sales outside of North America, Europe, or Japan	Sales Value
Global Sales	Global Sales	Sales Value

Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	total_sales
Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74	165.48
Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24	80.48
Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82	71.65
Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00	66.00
Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37	62.75

Video Games Sales Dataset | Kaggle

Scenario/Objective:

In this analysis, we will look at what characteristics and market conditions have led to greater video game sales. We will look at which market regions video game sales perform the best in. We will look at which video gaming companies and gaming consoles have had the largest impact on the industry and the best selling video games of all time. Finally, we will build Linear regression models to predict the future sales trend. We will use different columns as predictors analyze to see which sales region correlates higher with the overall sales trend and produce a more accurate model.

Data Manipulation:

```
vg.isnull().sum()
```

```
Rank            0
Name            0
Platform        0
Year          271
Genre           0
Publisher      58
NA_Sales        0
EU_Sales        0
JP_Sales        0
Other_Sales     0
Global_Sales    0
dtype: int64
```

```
vg2 = vg.loc[vg['Year'] <= 2016.0]
vg2.sort_values(by = ['Year'], ascending = False).head(5)
```

	Rank	Name	Platform	Year	Genre
13435	13437	Mighty No. 9	PS4	2016.0	Platform
13100	13102	Mirror's Edge Catalyst	PC	2016.0	Platform
10827	10829	Taiko no Tatsujin: Don Don! Mystery Adventure	3DS	2016.0	Action
4509	4511	Overwatch	PC	2016.0	Shooter
9680	9682	Lego Star Wars: The Force Awakens	PS3	2016.0	Action

```
vg2.isnull().sum()
```

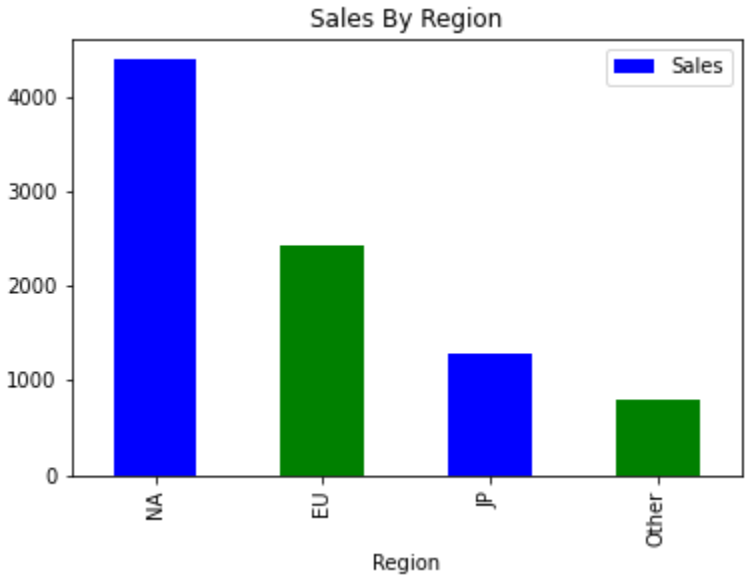
```
Rank            0
Name            0
Platform        0
Year            0
Genre           0
Publisher      36
NA_Sales        0
EU_Sales        0
JP_Sales        0
Other_Sales     0
Global_Sales    0
total_sales     0
dtype: int64
```

In order to filter out many of the missing values, games that were included in the data set from after the year 2016 were filtered out, as these games were only partially filled in.

Data Exploration:

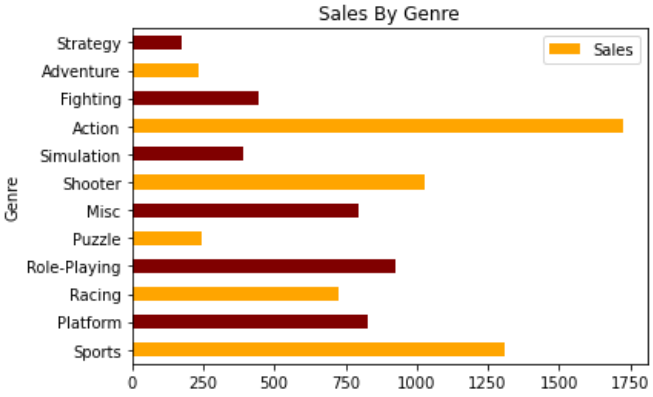
Regional Breakdown

As we can see from the histogram, North America has been the dominant market for video game sales throughout history. With $44 billion in sales, the North American market almost doubles the next best performing market, Europe at $24 billion. The North American market has accounted for almost 50% of all video game sales to date. As a whole, the North America and Europe markets make up a staggering 75% of all video game sales, as all other regions have only accounted for around $21 billion of the total $90 billion in video game sales to date.



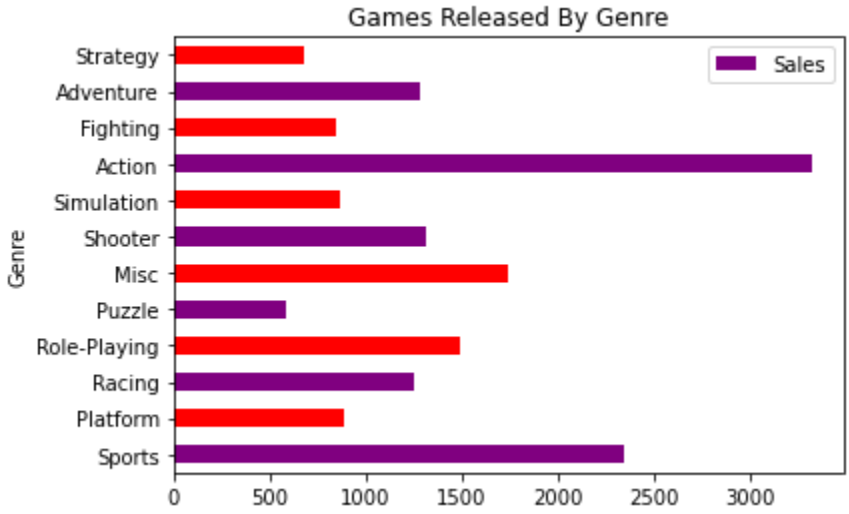
Analysis by Genre

The next step in the analysis is to examine which style of video games sell the best. To do this we will create a bar plot that shows the total sales from each video game genre. It is clear that Action and Sports are the most popular video game genres. These two



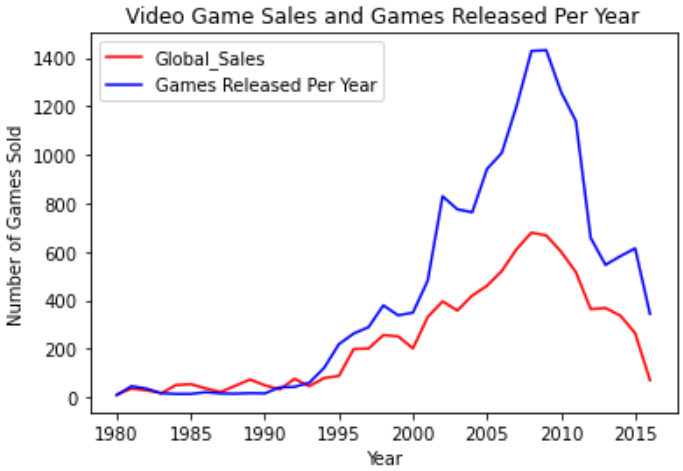
categories make up 34% of total sales. The remaining 10 categories all make up 10% or less of the total sales. From analyzing the total sales by genre in this data set, it is clear that Sports and Action games are significantly favored over any other genre.

This is a similar bar plot that represents the amount of video games released in each genre. Looking at this graph we can confirm what we found in the previous graph. While some genres have discrepancies between sales and releases, the Action and Sports genres still severely outnumber their competitors.



Sales Trend

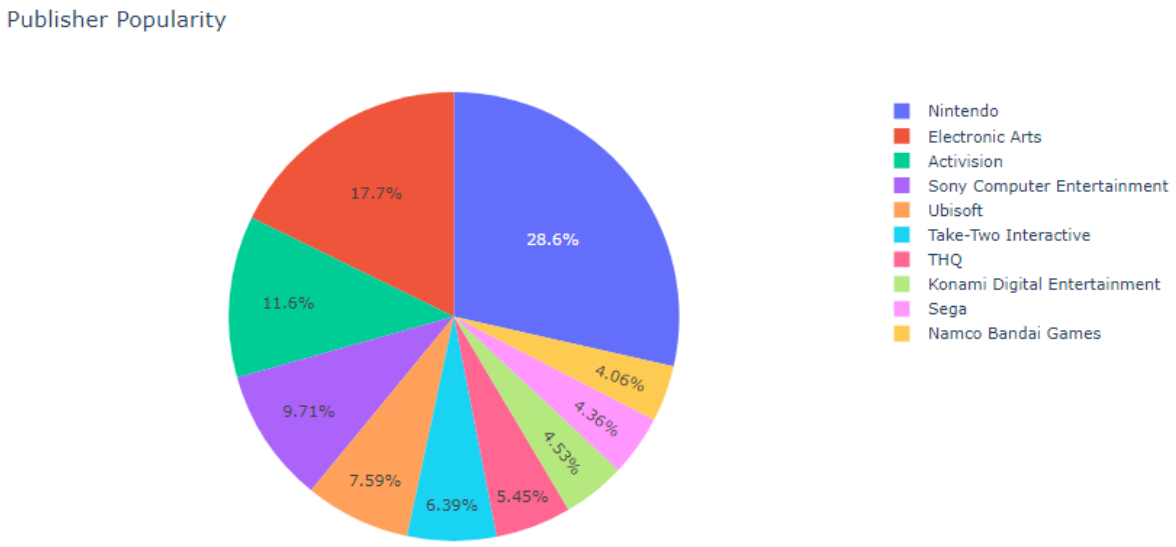
The first decade (1980-1990) of the video gaming industry saw mediocre results as the market for games did not grow very much at all and the same amount of video games were released year to year. However once the 1990s began, the explosion of video game popularity was ushered in as the video game sales trend began to start its 20 year long run of growth. From the 1990s through the early 2010s,



video games became one of the premier forms of entertainment in the world, and the amount of video games sold per year began to skyrocket. From 1990 to 2010, the number of video games sold per year increased 115% from 98 to 1200 games released per year. The decline in sales and games released per year during the 2010s was a result of the change in the way people consumed content, as well as a change in the business model of gaming companies. As streaming took over as the primary way to view content, gaming companies began to release less games and rely more on things like in-game purchases and updated gameplay.

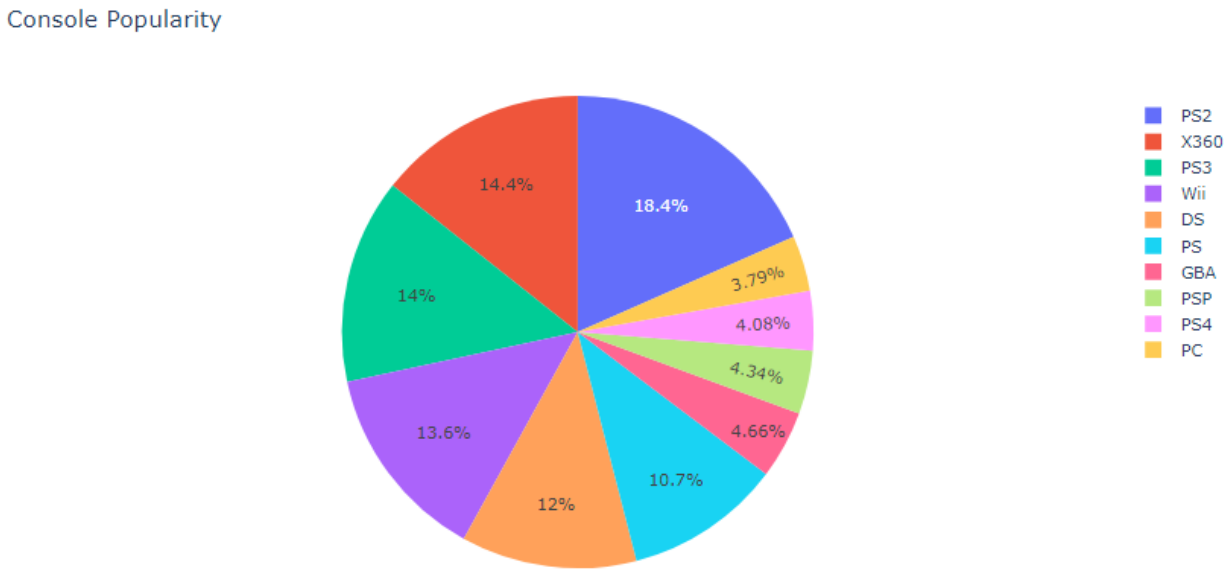
Game Publishers

In the breakdown of the top 10 video game publishers of all time we see that Nintendo has accounted for over a quarter of all worldwide video game sales. Electronic Arts, a popular publisher known for the EA Sports game line, and Activisioncreater of the popular Call of Duty game series also make up significant portions of all time video game sales. Seeing



Console Popularity

Our next step in the analysis is to examine which consoles the vast majority of gamers are playing on. Here we see a pie chart representing the top 10 most popular gaming consoles based on games sold. We see here the influence that popular gaming companies like Nintendo, Sony, and Microsoft have on the gaming industry. Sony has created 5 of the top 10 gaming consoles of all time with their PlayStation line (PS1, PS2, PS3, PS4, PSP).

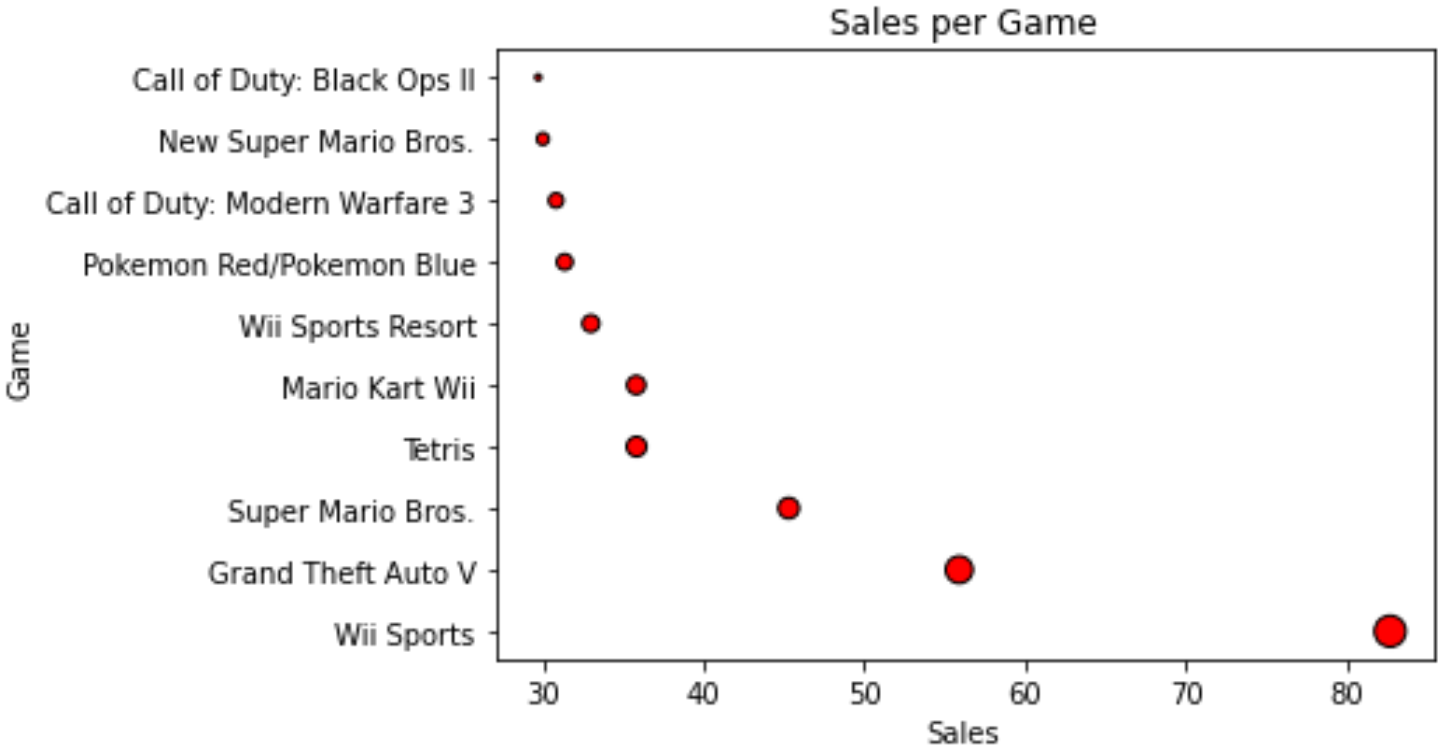


Nintendo has three consoles in the top 10 (Wii, DS, Gameboy Advanced) while Microsoft's Xbox 360 holds the title of 2nd most popular console of all time. These three companies have created the top 9 most popular gaming consoles of all time.

Individual Game Sales

In this scatter plot, we see the top 10 selling video games of all time. As one would assume, all

of the games in the top 10 are universally recognizable. Just like in our analysis of gaming consoles, Nintendo is among the most successful video game publishers when it comes to making games as well. Nintendo accounts for 6 of the top 10 selling video games of all time. Another very popular line of games created by another well known publisher *Activision* is the Call Of duty line of games. Activision released 2 Call of Duty games that are members of the top 10 selling video games of all time. We can conclude that Wii Sports is the undisputed most popular video game of all time, selling almost a third more than the second most popular game ever, Grand Theft Auto V.



Models

Linear Regression Models

```
X1 = vg2.loc[:, ["NA_Sales","EU_Sales"]]
y1 = vg2['Global_Sales']
```

```
#  The Mean absolute error value represents the average value of error,
#  or the difference between the actual values and predicted values.

from sklearn.metrics import mean_absolute_error
print('MAE:', mean_absolute_error(y_test,y_pred))
```

```
MAE: 0.11310932400920175
```

```
#The Mean square error represents the average of the squares, of each error

from sklearn.metrics import mean_squared_error
print("MSE",mean_squared_error(y_test,y_pred))
```

```
MSE 0.08995301255905441
```

```
#Square root of the mean square error

import numpy as np
print("RMSE",np.sqrt(mean_squared_error(y_test,y_pred)))
```

```
RMSE 0.29992167737436787
```

```
#the r squared value represents the overall accuracy of the model. Model 1 is 95% accurate.

from sklearn.metrics import r2_score
r2 = r2_score(y_test,y_pred)
print(r2)
```

```
0.9543806090362525
```

\

The first regression model was built to predict the global_sales column using the North American sales and European Sales and a 30% test size. The model came back very accurate with a MAE, MSE, and RMSE all below 1. The overall r squared accuracy of the model was 95% making this a very effective prediction model. As the North American and European sales are highly correlated with the total global sales, these two markets will continue to drive growth in the video game industry.

Model 2

```
X = vg2.loc[:, ["JP_Sales","Other_Sales"]]
y = vg2['Global_Sales']
```

```
from sklearn.metrics import mean_absolute_error
print('MAE:', mean_absolute_error(y_test2,y_pred2))
```

```
MAE: 0.27206001513369454
```

```
from sklearn.metrics import mean_squared_error
print("MSE",mean_squared_error(y_test2,y_pred2))
```

```
MSE 0.6119906185221453
```

```
import numpy as np
print("RMSE",np.sqrt(mean_squared_error(y_test2,y_pred2)))
```

```
RMSE 0.7822982925471238
```

```
from sklearn.metrics import r2_score
r2 = r2_score(y_test,y_pred2)
print(r2)
```

```
0.6896308583975583
```

In our second linear regression model we see that the result is much less accurate. In this model we use the JP_Sales and Other_Sales column as predictors. These columns represent the sales of all other regions outside of North America, and Europe. While the MAE, MSE, and RMSE are all relatively low, the r squared value gives the model an accuracy of 68%.

Conclusion

From this analysis we get a closer look into the video game industry and the different conditions that lead to greater sales. From our initial analysis, we can conclude that the video game market

is a global, widespread market, yet very specialized. We see that geographically , the North American market is the main driver for the industry. We also see that by looking at the production and sales by genre, console, and publisher that the industry is dominated by the popularity of only a few genres and publishers. When analyzing the results of our linear regression models, we see that the North American and European video game markets have a much larger correlation with the overall sales trend, than the other regions throughout the world. We expect these two regions to continue to drive the growth for video game popularity in the future based on our predictions.