Analysis of Console Video Game Sales

Robert Johnson

1. INTRODUCTION

Video gaming has come a long way since the early days of Pong and Pac-man. The aggressive competition between gaming companies has driven the technological leaps to superior features and video gaming quality we see today. The more than 70 gaming consoles have produced 100's of memorable characters that are embedded in international culture.

In 2019, video game sales were more than $116 billion globally, and nearly two-thirds of American homes have household members who play video games regularly.

The game industry is also often at the forefront of computer technology and drives popular culture with console and PC "gamers" make up the bulk of sales.

This study will focus on gaming consoles and use that data to predict future game sales.

1.1 DATA SET

Gregory Smith used a web scraper on VGChartz Video Games Sales to make the foundation of this information, and the data set is extended with another web scrape from Metacritic. This data set has about 18,000 entries.

The data includes 17 metrics items: Name, Platform, Year_of_Release, Genre, Publisher, NA_Sales, EU_Sales, JP_Sales, Other_Sales, Global_Sales, Critic_score, Criticcount, Criticscore, User_score, User_count, Developer, Rating.

https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings

1.2 DATA IMPORTATION AND CLEANING

Next, the game sales dataset was imported as a .csv file, which needed to be converted into a panda data frame.

All NA's and six columns: Critic_score, Criticcount, Criticscore, User_score, User_count,

Developer, Rating were removed for the data set. Next, the Year_of_Release object was converted to an integer.

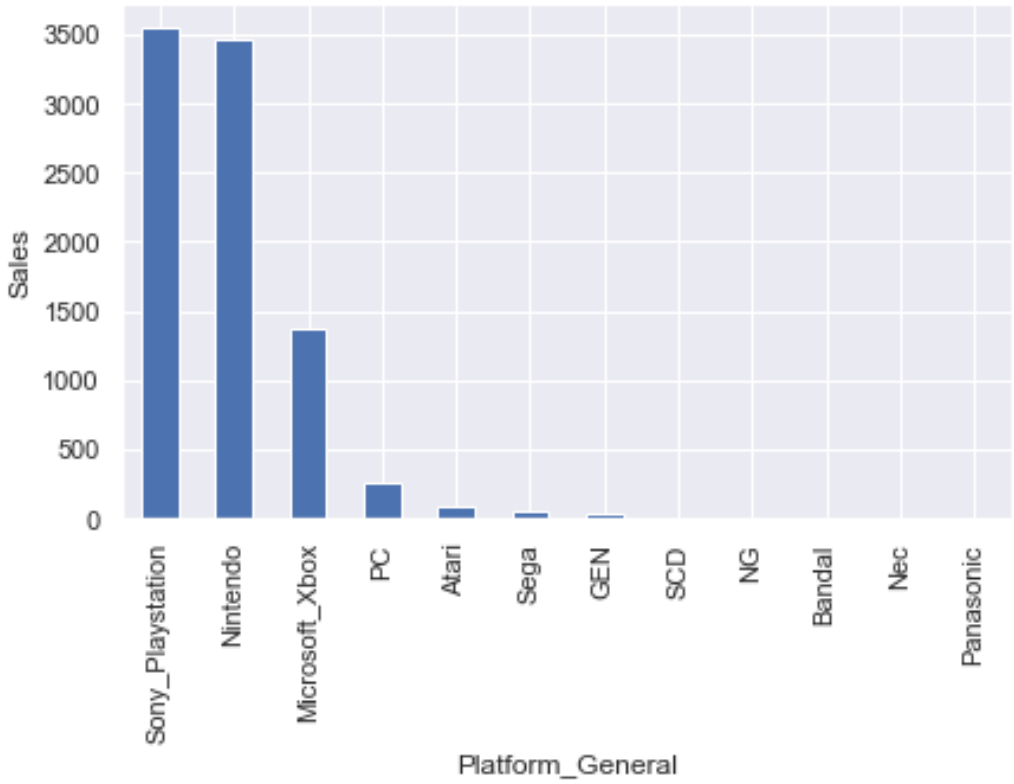
The names of the platform names were simplified and combined the different versions of the platforms.

The end data set was 16,000+ row and 11 columns, which this study used to explore the trends of the industry and attempt to build a model that will predict future global sales.

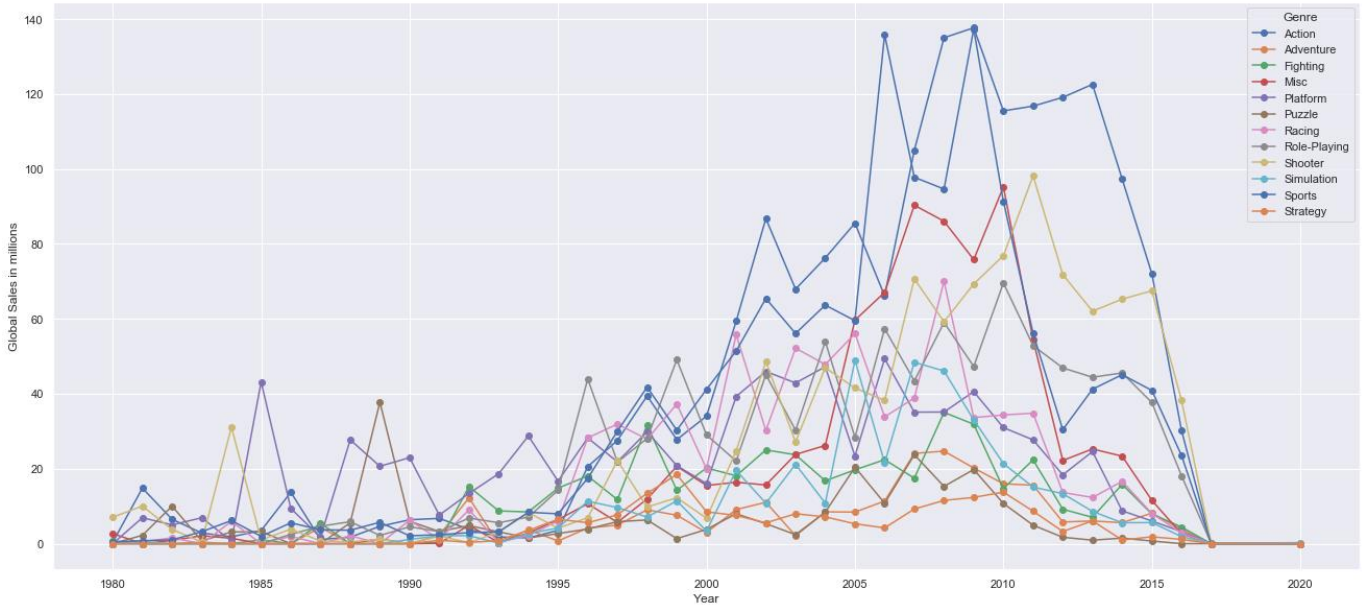
```
class 'pandas.core.frame.DataFrame'>
Int64Index: 16416 entries, 0 to 16718
Data columns (total 11 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Name              16416 non-null  object
 1   Platform          16416 non-null  object
 2   Year_of_Release   16416 non-null  int32
 3   Genre             16416 non-null  object
 4   Publisher         16416 non-null  object
 5   NA_Sales          16416 non-null  float64
 6   EU_Sales          16416 non-null  float64
 7   JP_Sales          16416 non-null  float64
 8   Other_Sales       16416 non-null  float64
 9   Global_Sales      16416 non-null  float64
 10  Platform_General  16416 non-null  object
dtypes: float64(5), int32(1), object(5)
```

2. DATA EXPLORATION

Initially, the relationship between variables in game sales data was explored by looking at global sales across all platforms from 1980 to 2019. By combining the different versions of the platform into one variable, the sales distribution per platform was easier to categorize.



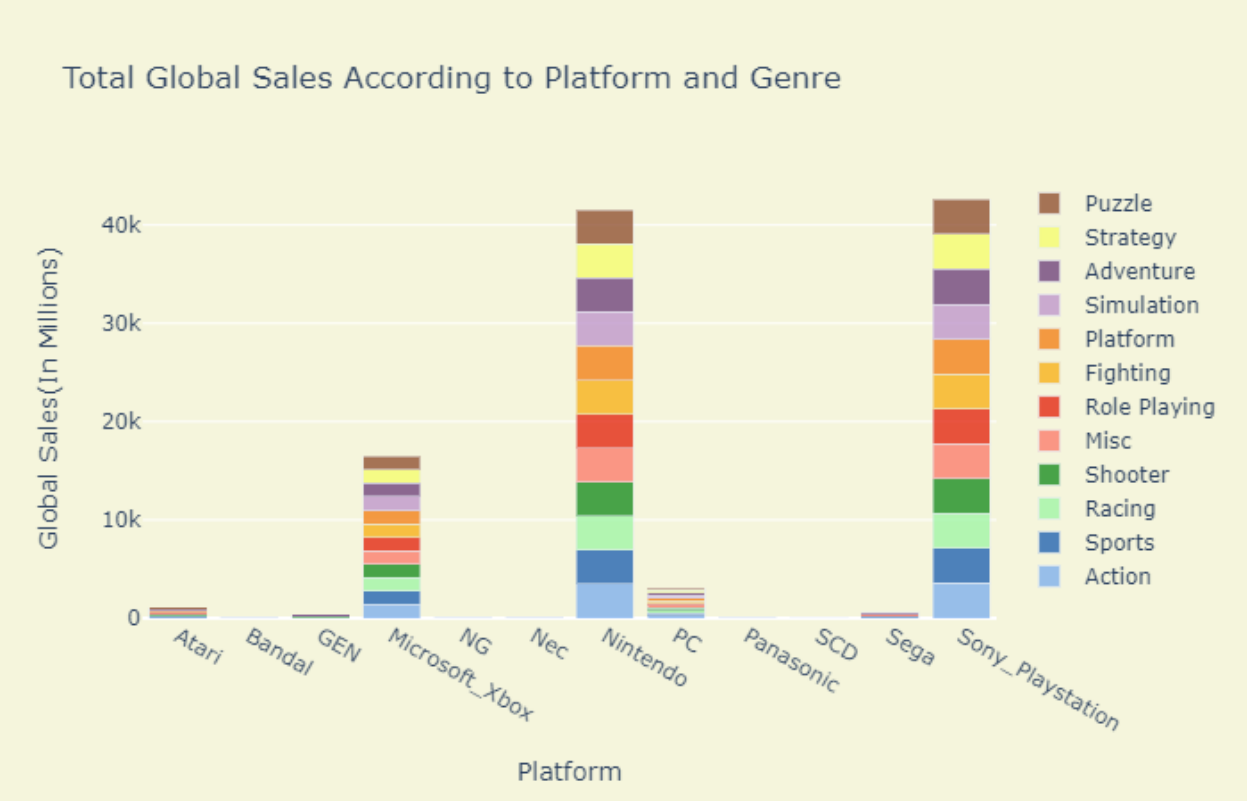
The global sales data has a right skew with is increasing over time with a significant jump that happens soon after the turn of the century.



The graph above also points to the importance of genre concerning total sales; the action and sports genre make up most of the sales for most years.

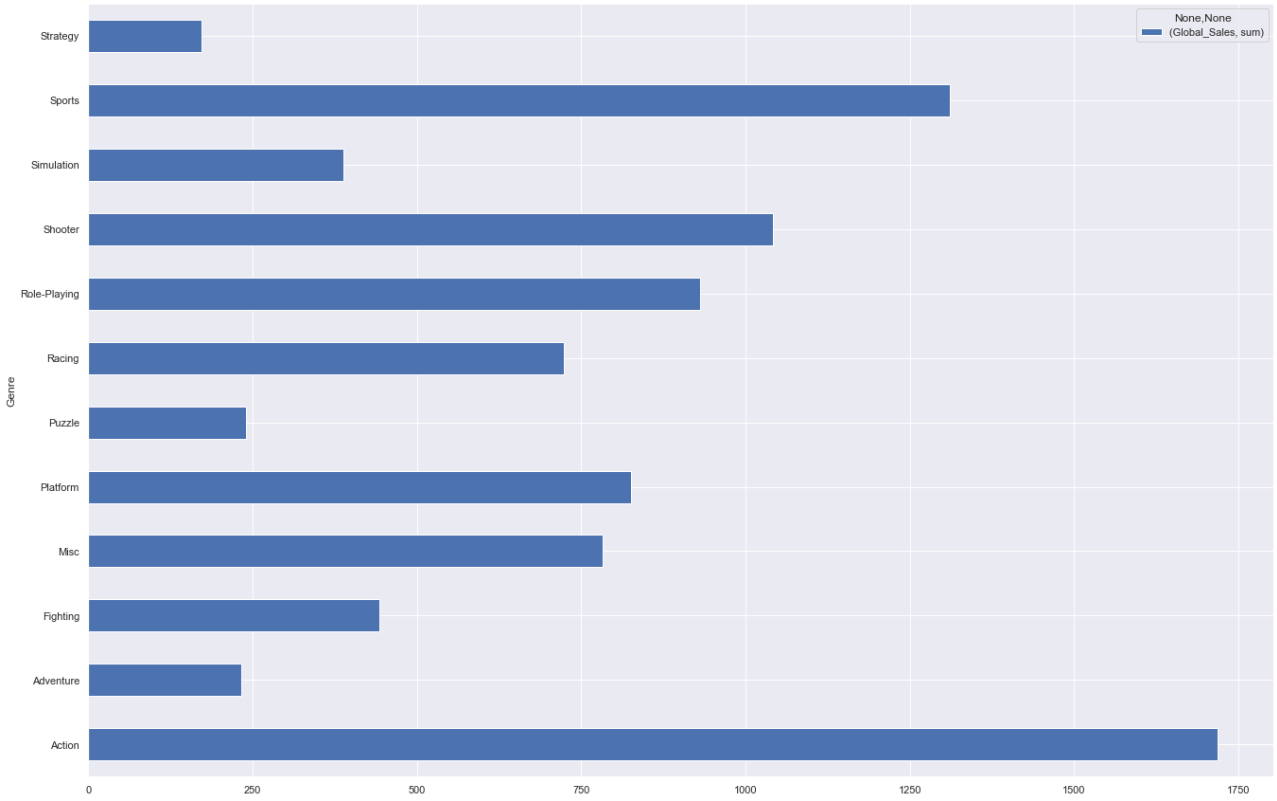
3. PATTERN OF DATA

One of the most striking information the game sales data showed is that there are three leading platforms for game sales Sony PlayStation, Nintendo, and Xbox. Microsoft, the number three platform, has more sales than the next nine platforms combined.



Nevertheless, most of the calculations use all 12 of the platforms' information.

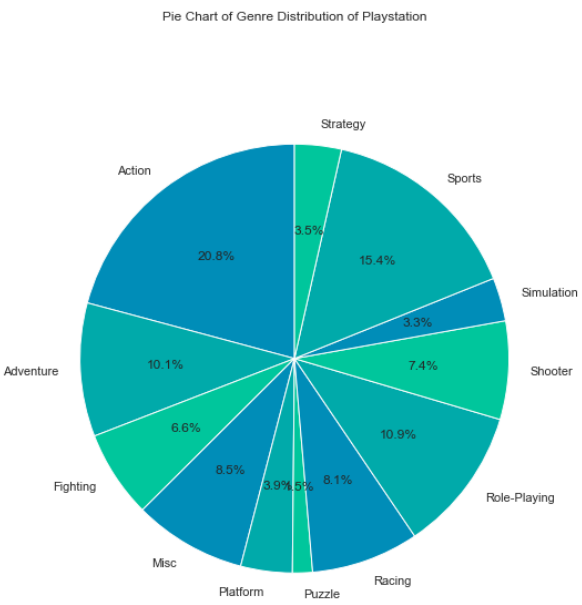
Evaluating game sales by genre is an effective way to sort game sales information. The precise delineation of sales by gene helped identify game sales for each genre across all platforms.

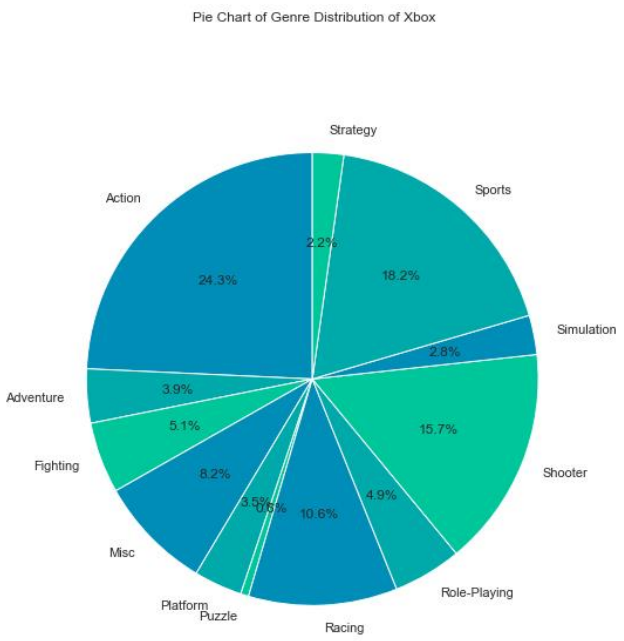


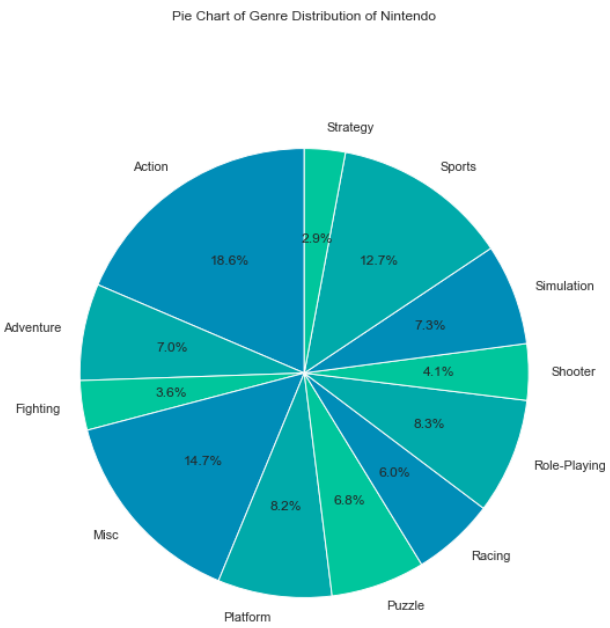
This data set, the best-selling game for each platform with total global sales, was able to be identified. The ability to drill down to a specific game, its genre, and sales in the year of the release highlight the usefulness of this information.

The best-selling game for each genre across the history of platform gaming was also identified.

As shown in the bar graph above, the action genre has the most sales of any genre, and this is consistent with the finding when global sales are broken down by percentages of sales by platform.

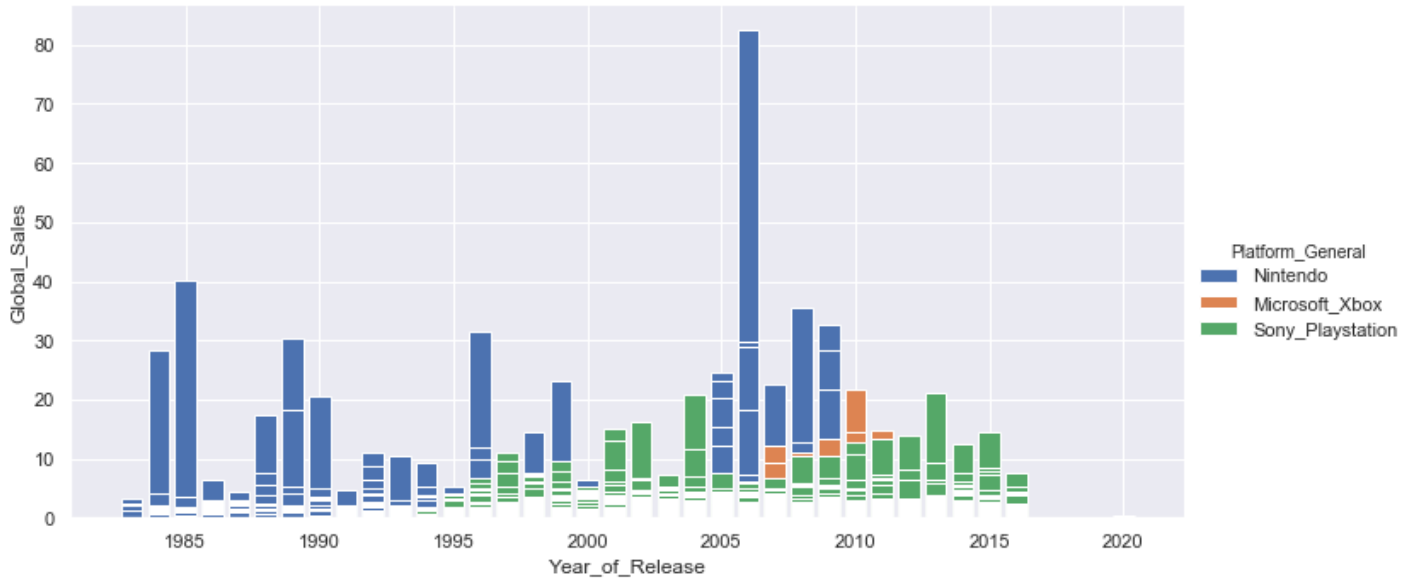






The genre distribution for each console gives insight into the type of games the platform users prefer. All three of the major platforms have action as the largest percentage. PlayStation and Xbox have sports second-largest percentage of sales, Nintendo has miscellaneous. The third-place sales are different for each platform PlayStation has role-playing games, Nintendo has sports, and Xbox has shooters.

The last and most telling pattern was year-over-year sales. From the 1980s to the 2000s, Nintendo has led all game sales, except for one year. From 2000 on, PlayStation has led except for five-year lead by Nintendo and two years led by Xbox.



4. PREDICTIVE MODELS

The original goal of this study was to use this data to predict future game sales. This study used several models, including linear regression and polynomial regression, to predict future game sales in general, by the company or by genre. None of the prediction models product an R-squared above .15 most of them were below .05 for both the train and test set of data.

5. FINDINGS

As previously stated, predictions were not possible with this data set; however, the historical data in this data set gives a fascinating snapshot of this industry.

Nintendo platforms account for the most video game sales overall, and nine of the twelve genres top-selling games are exclusive to Nintendo's platform.

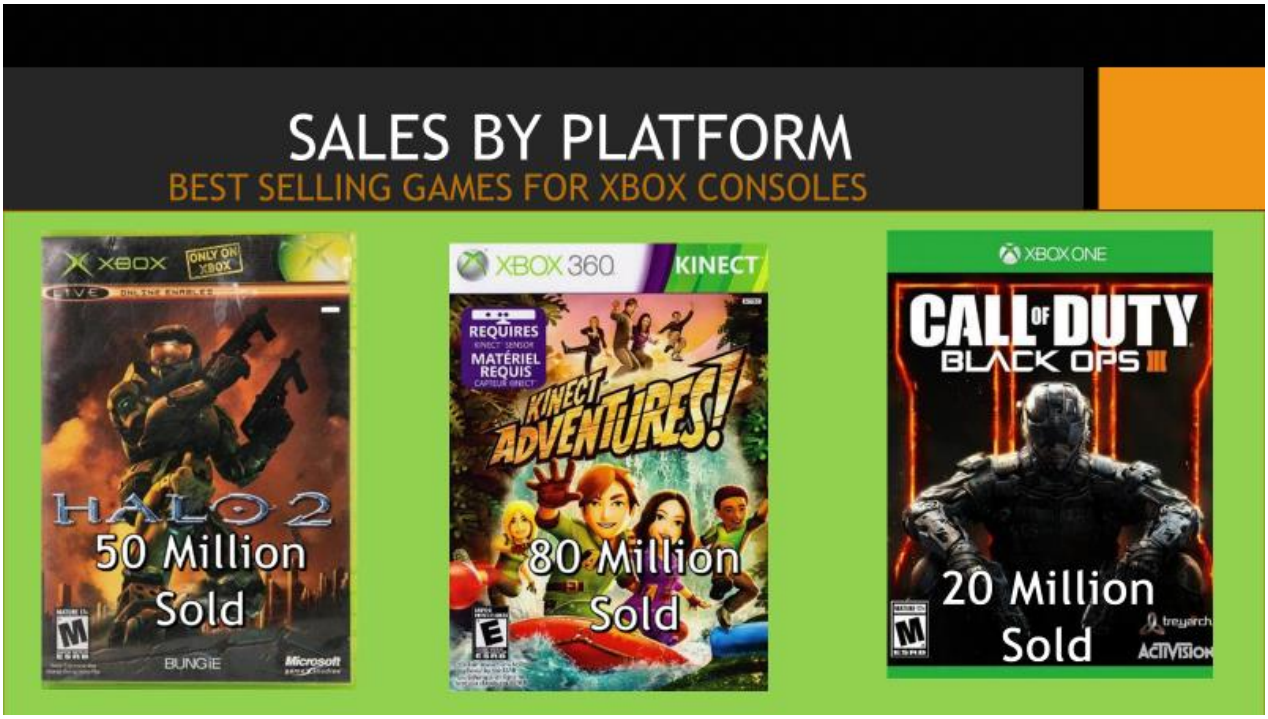
Starting in 2012, PlayStation has should more game than Nintendo and Xbox.

The top-selling games on each platform show the diverse taste of their customer base and can guide publishers and game makers on the type of games that might appeal to users.

PlayStation top games:

PlayStation has released six consoles, both handheld, and home-based. Above is the top game for each console. PlayStation's top-selling games show that its gamers prefer M-rated, over 18, fast pasted games with shooting and online play capability. The three Grand theft auto games and Call of Duty were also games that released when PlayStation lead Nintendo and Xbox in-game sales.

Xbox top games:



Xbox has released three home consoles. Above is the top game for each Xbox consoles. PlayStation and Xbox are chasing the same customers. They had the same top-selling game in the last year observed. Xbox customers also prefer M-rated, over 18, fast-paced shooters with online play. Xbox is a distant third in sales to PlayStation and Nintendo.

Nintendo games:

Nintendo has released six consoles and handheld devices. A majority of Nintendo's top games are E-rated, everyone, kid-friendly. Nintendo has the overall sales lead because its focus is on children and casual gamer, which has no significant competitor. From a console position, they are in the best place to differentiate themselves and experience exponential growth.

5. FUTURE PROJECTS

To predict future game sales would take more statistical and categorical information, then is in this data set. Finding this is information would be a worthy pursuit. The information needed would include the sales number of previous versions of the game, the release date of the prior version of the game if it a sequel, the player rating of the past or similar games, the reputation of the publisher, and the genre of the game.

The growth of mobile games makes this entirely different sector an exciting subject. These sales are only digital, unlike consoles, which mean companies have a complete record of the games sold. However, mobile game publishers are only open with their revenue information with their investors.