VIDEO GAME SALES

https://wall.alphacoders.com/big.php?i=789452

Video Game Sales

Example of Street Fighter:
“EVO Moment #37”

https://youtu.be/JzS96auqau0?t=5

Video Game Sales

Multi-Billion Dollar Industry

http://thoughtforyourpenny.com/technology/video-games/10-best-gaming-podcasts-to-listen-to-in-2019/
https://oremc.com/o-raining-money-facebook#iLightbox[postimages]/0

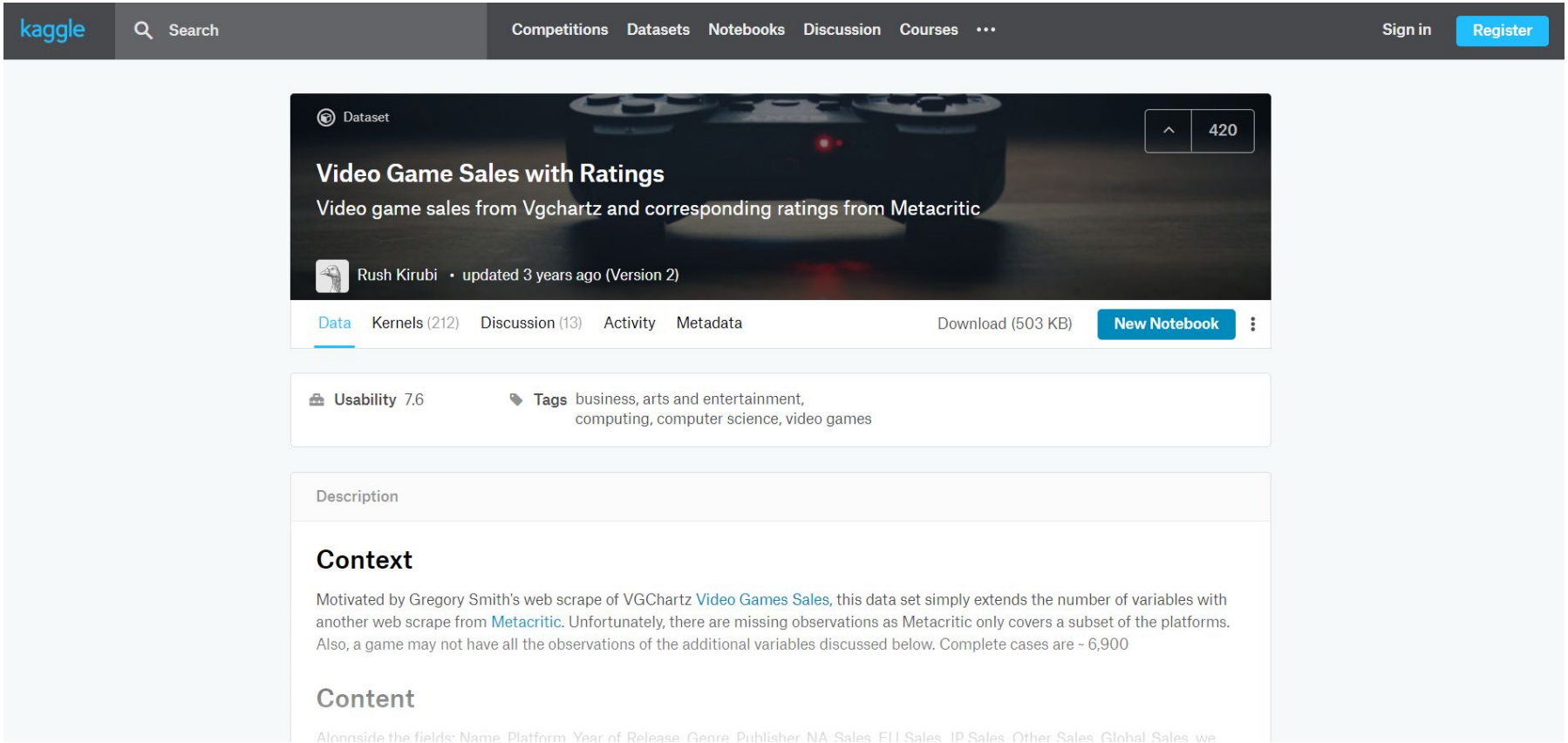
Racing to the Objectives

- Observe relationship between video game scores and sales
- Other features of a high selling video game
- Machine learning model to identify high selling games

https://wccftech.com/review/onrush-beyond-finish-line/

https://www.flickr.com/photos/146274265@N03/46559151605/

The Dataset



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

The Dataset

THE LEGEND OF ZELDA: BREATH OF THE WILD Switch

Nintendo | Release Date: Mar 3, 2017 | Also On: Wii U

Summary | Critic Reviews | User Reviews | Details & Credits | Trailers & Videos

97 Metascore
Universal acclaim
based on **109 Critics**

What's this?

Summary: Forget everything you know about The Legend of Zelda games. Step into a world of discovery, exploration and adventure in The Legend of Zelda: Breath of the Wild, a boundary-breaking new game in the acclaimed series. Travel across fields, through forests and to mountain peaks as you discover... **Expand**

8.5 User Score
Generally favorable reviews
based on **12710 Ratings**

Your Score 0

Developer: Nintendo
Genre(s): Action Adventure, Open-World
# of players: No Online Multiplayer
Cheats: On GameFAQs
Rating: E10+
More Details and Credits »

AWARDS & RANKINGS

- #1 Best Switch Game of 2017
- #1 Most Discussed Switch Game of 2017
- #1 Most Shared Switch Game of 2017

ESSENTIAL LINKS

- New Switch Releases by Date
- New Switch Releases by Title
- Upcoming Switch Release Calendar
- All-Time High Scores - Switch
- 2019 High Scores - Switch
- Hardware Reviews: Nintendo Switch

https://www.metacritic.com/game/switch/the-legend-of-zelda-breath-of-the-wild

The Dataset

Problems

- Duplicate video game names
- Games that were not released yet
- Did not have both Critic Score and User Score

Solutions

- Group dataset by video game names
- Video games before 2016
- Only games with both Critic and User scores

A	Name	Name of the game
A	Platform	Console on which the game is running
#	Year_of_Release	Year of the game released
A	Genre	Game's category
A	Publisher	Publisher
#	NA_Sales	Game sales in North America (in millions of units)
#	EU_Sales	Game sales in the European Union (in millions of units)
#	JP_Sales	Game sales in Japan (in millions of units)
#	Other_Sales	Game sales in the rest of the world, i.e. Africa, Asia excluding Japan, Australia, Europe excluding the E.U. and South America (in millions of units)
#	Global_Sales	Total sales in the world (in millions of units)
#	Critic_Score	Aggregate score compiled by Metacritic staff
#	Critic_Count	The number of critics used in coming up with the Critic_score
#	User_Score	Score by Metacritic's subscribers
#	User_Count	Number of users who gave the user_score
A	Developer	Party responsible for creating the game
A	Rating	The ESRB ratings (E.g. Everyone, Teen, Adults Only..etc)

The Dataset

Columns for analysis

- Name
- Year of Release
- Genre
- Publisher
- Global Sales
- Critic Score
- User Score

A	Name	Name of the game
A	~~Platform~~	Console on which the game is running
#	Year_of_Release	Year of the game released
A	Genre	Game's category
A	Publisher	Publisher
#	~~NA_Sales~~	Game sales in North America (in millions of units)
#	~~EU_Sales~~	Game sales in the European Union (in millions of units)
#	~~JP_Sales~~	Game sales in Japan (in millions of units)
#	~~Other_Sales~~	Game sales in the rest of the world, i.e. Africa, Asia excluding Japan, Australia, Europe excluding the E.U. and South America (in millions of units)
#	Global_Sales	Total sales in the world (in millions of units)
#	Critic_Score	Aggregate score compiled by Metacritic staff
#	~~Critic_Count~~	The number of critics used in coming up with the Critic_score
#	User_Score	Score by Metacritic's subscribers
#	~~User_Count~~	Number of users who gave the user_score
A	~~Developer~~	Party responsible for creating the game
A	~~Rating~~	The ESRB ratings (E.g. Everyone, Teen, Adults Only..etc)

Critic Score vs User Score

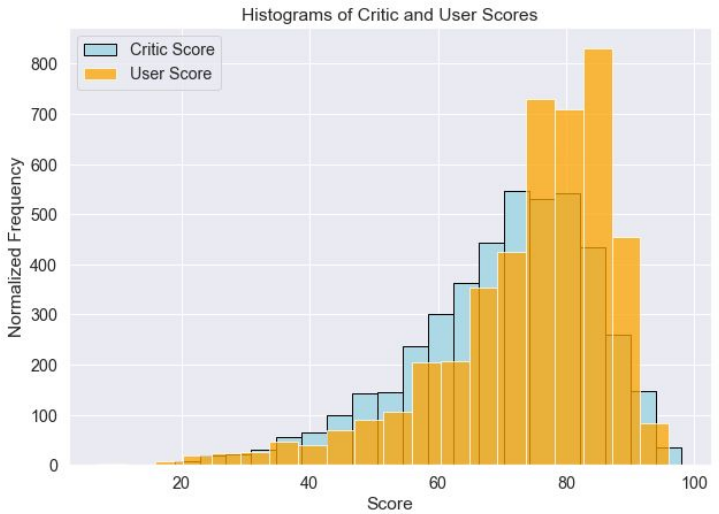
Are they different?

https://www.hawkinspersonnel.com/cheerful-business-people-showing-thumbs-up-isolated-on-white/

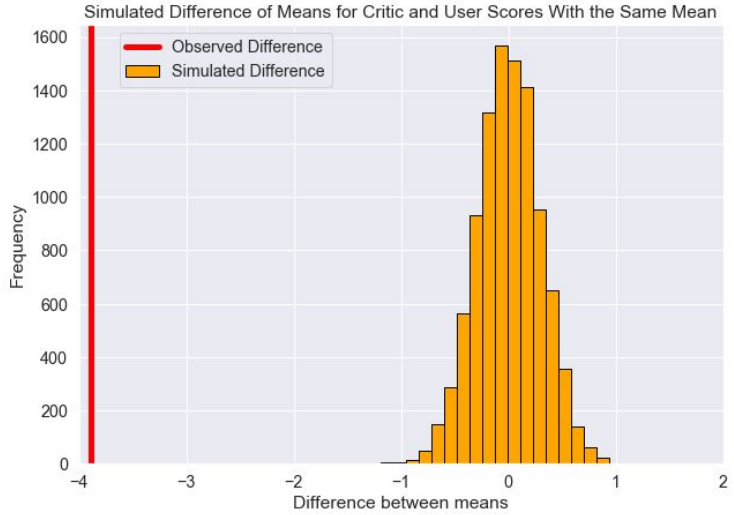
https://striveliving.ca/adults-services/strive-centre/group-of-people-showing-thumbs-up/

Critic Score vs User Score

- Mean of Critic Score: 70.45
- Mean of User Score: 74.34
- Difference of means: 3.88

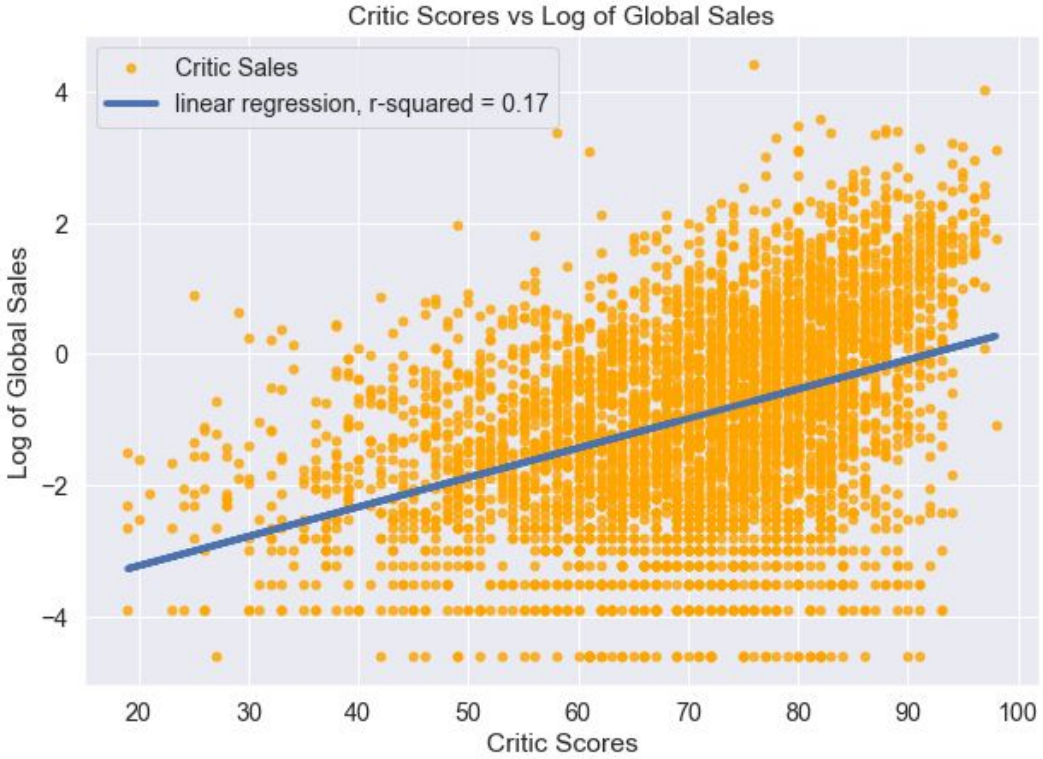


- Bootstrap resampled a simulated dataset where mean of Critic score = mean of User score
- Sampling distribution of difference of simulated means
- P-value = 0.00... for observed difference



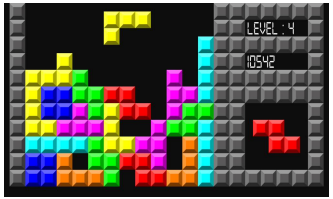
Critic Score and Global Sales

- Scatter plot of Global Sales and Critic Scores
- Majority of data points below 20 million units
- Log transformation of Global Sales
- Simple linear regression
- R-squared value of 17%
- Using Machine Learning to fit additional features
- Produced new R-squared of 30%



Video Game Genres

Strategy



Puzzle

Adventure

Action

Role-Playing

Simulator

Fighting

Racing

Platform

Sports

Misc

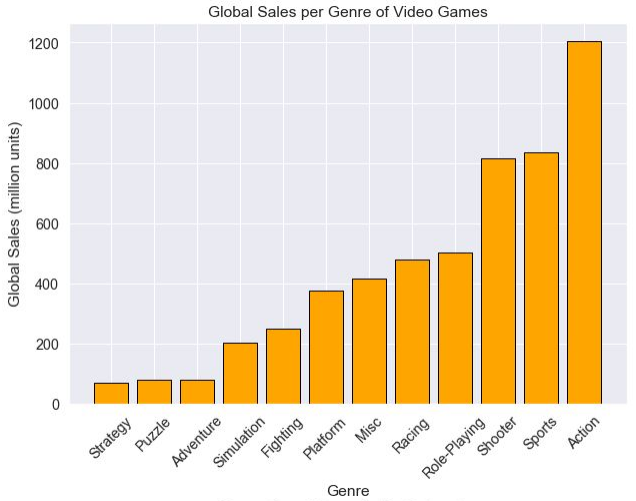
Shooter

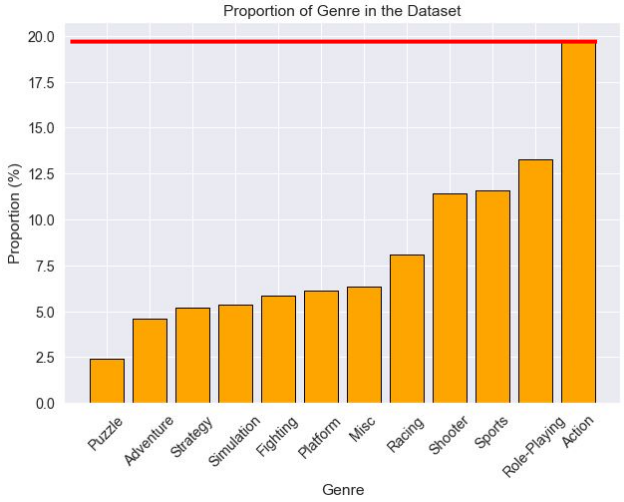
**See notes for links to images

Genre vs Global Sales

- Total global sales per Genre
- Action games have the highest total global sales
- About 400 million units higher than Shooters or Sports.

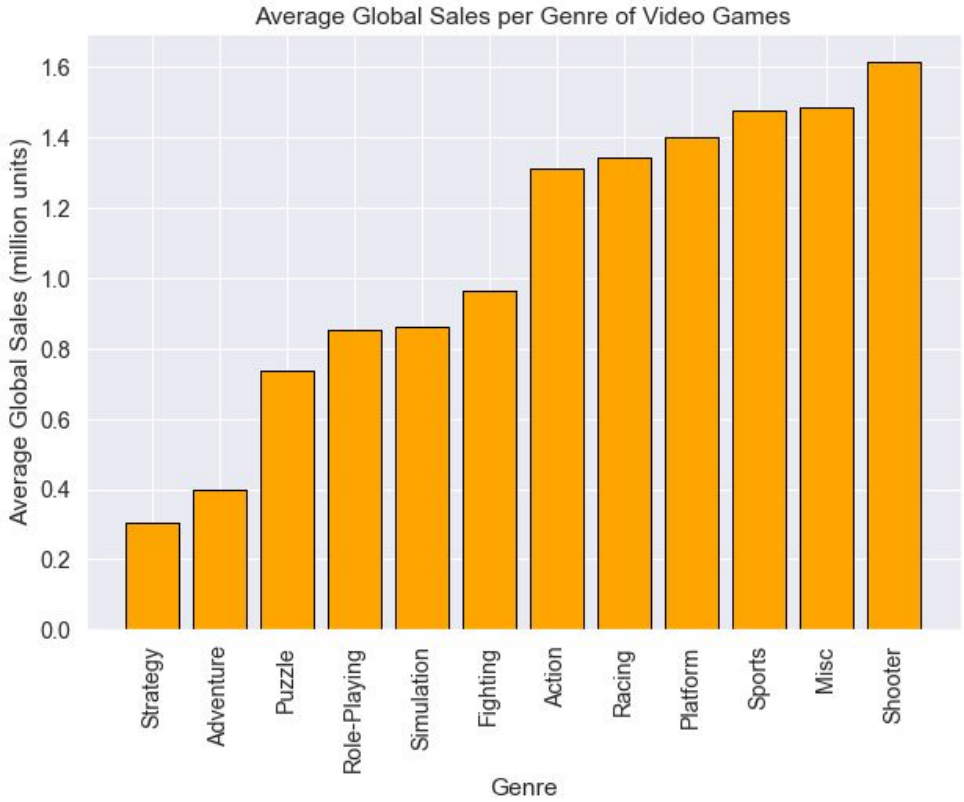
- 20% of all video games were action
- More action games = more total sales
- What is a better way to see genre and global sales?
- Look at average sales!





Genre vs Average Global Sales

- Average global sales of each genre
- Action games have moved to 6th place
- Shooters have highest average sales at about 1.6 million units

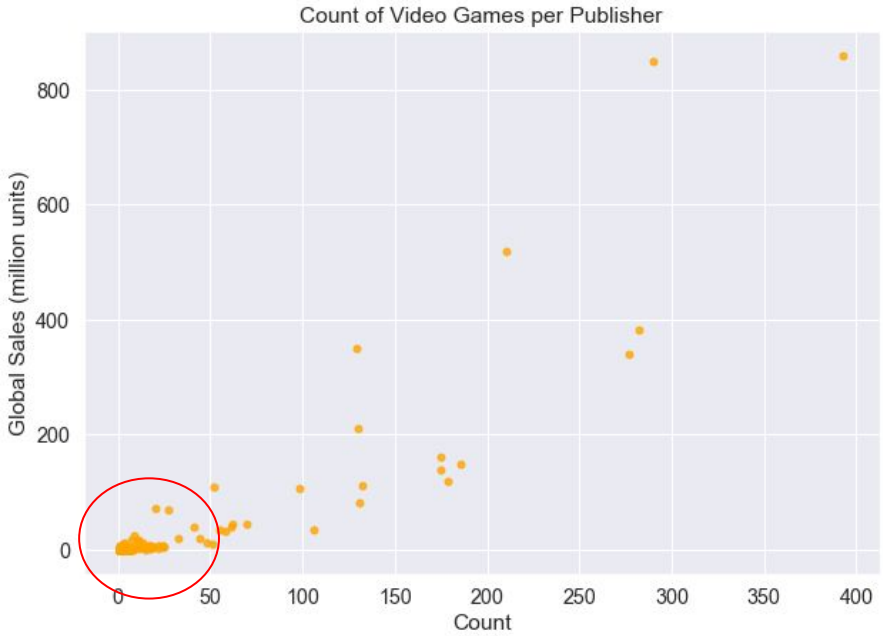


Publishers

https://www.gamedesigning.org/game-development-studios/

Publisher Game Count vs Global Sales

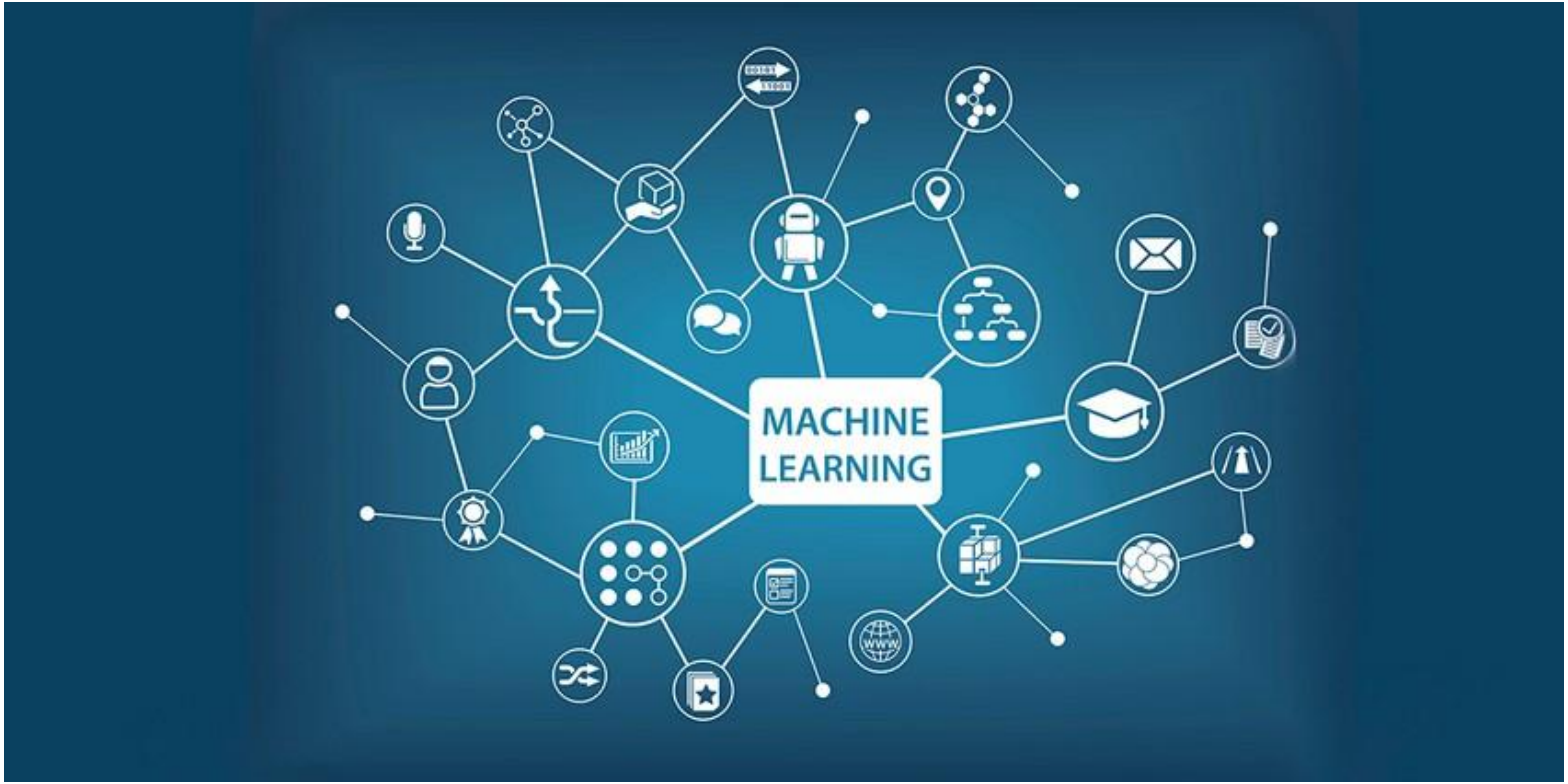
- Count of games per publisher
- The more games you make the more money you make!
- Many data points at lower left
- Log transformation of count and global sales



Publisher Game Count vs Global Sales

- Log transformation of count and global sales
- Fitting a linear regression line
- R-squared of 68%

Machine Learning



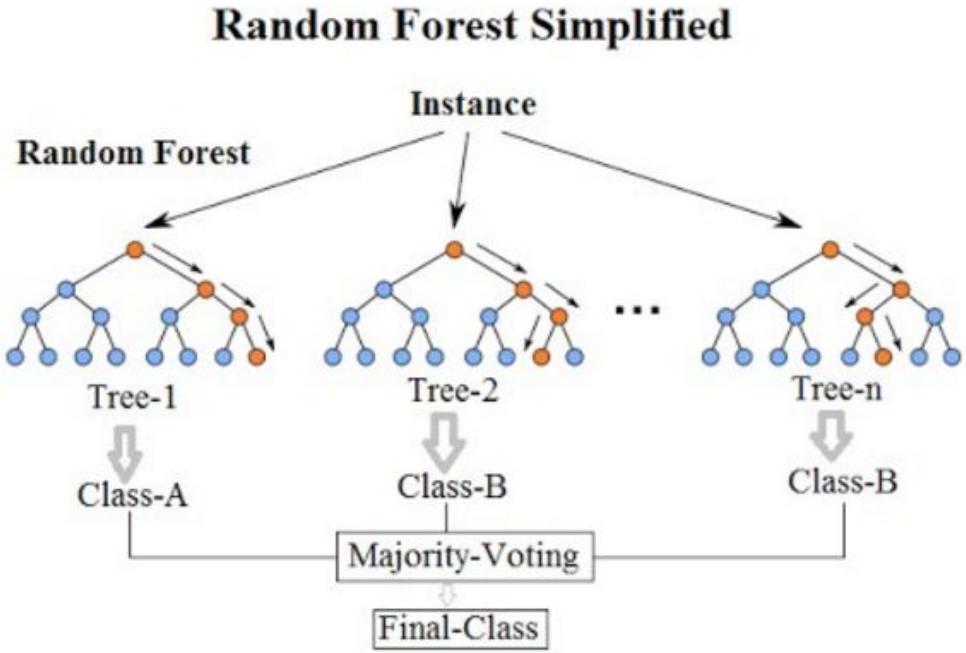
https://towardsdatascience.com/machine-learning-for-beginners-d247a9420dab

Classification

- Supervised learning for predictions
- Need labeled data
- Range of global sales values is 0.01-82 million units
- High selling = greater than 1 million units
- Video games label as either high selling or not

Random Forest Classifier

- Ensemble learning method
- Majority vote of many decision trees



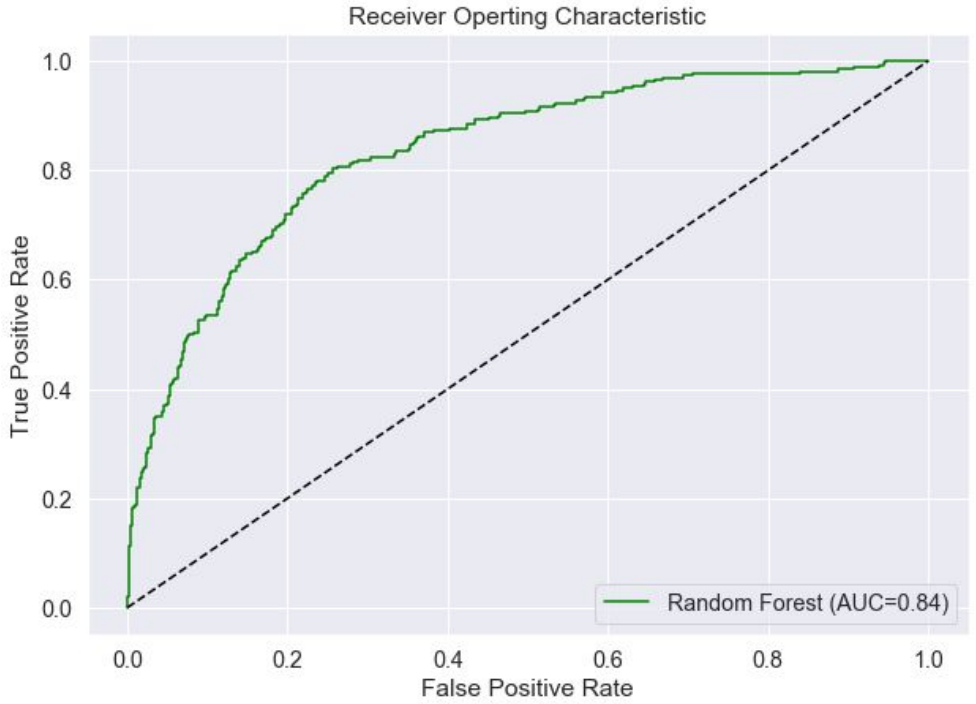
https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d

Steps Taken for Training and Testing

- Split the labeled dataset into features variable and a target variable (the label)
- Encoded categorical features and the target variable
- Split the dataset into training set and a hold-out set
- Tuned the hyperparameters of the random forest using sklearn GridSearchCV with the training set
- Tested random forest model by fitting on training set and testing on hold-out set

Random Forest Classifier Results

- Accuracy = 79%
- Precision = 73%
- AUC = 84%



Key Takeaways

- Higher critic scores tend to have higher global sales
  - Critic score vs log global sales
  - R-squared of 17%
- Shooters had a higher average global sales
  - Shooter average global sales: ~ 1.6m units
  - Action average global sales: ~1.3m units
  - 20% of all video games were action
- Publishers that make more games make more money
  - Log of count vs log of global sales
  - R-squared of 68%
- Random Forest Classifier for Prediction
  - Accuracy of 79%
  - Precision of 73%
  - AUC of 84%

https://vsbattles.fandom.com/wiki/Sora_(Kingdom_Hearts)

Other Avenues to Explore and Areas to Improve

- Look into User Scores
- Better data cleaning and wrangling
- Neural Network

https://www.geek.com/games/11-ways-video-games-can-make-you-a-better-person-1639865/

https://www.shutterstock.com/video/clip-32553706-retro-game-style-infinite-tunnel-flight-seamless