# Video Game Sales

Prediction modeling with
critic and user ratings

Laura Howard

# What video game features best predict unit sales?

- User Ratings
- Critic Ratings
- Developer
- Platform/Device
- Genre
- Year released


**Business applications:**

- Can developers predict how well their next game will sell given a set of user and critic ratings from a small group of pre-release testers?
- Can investors predict success of upcoming games and buy/sell development company stock accordingly?

# Data Acquisition

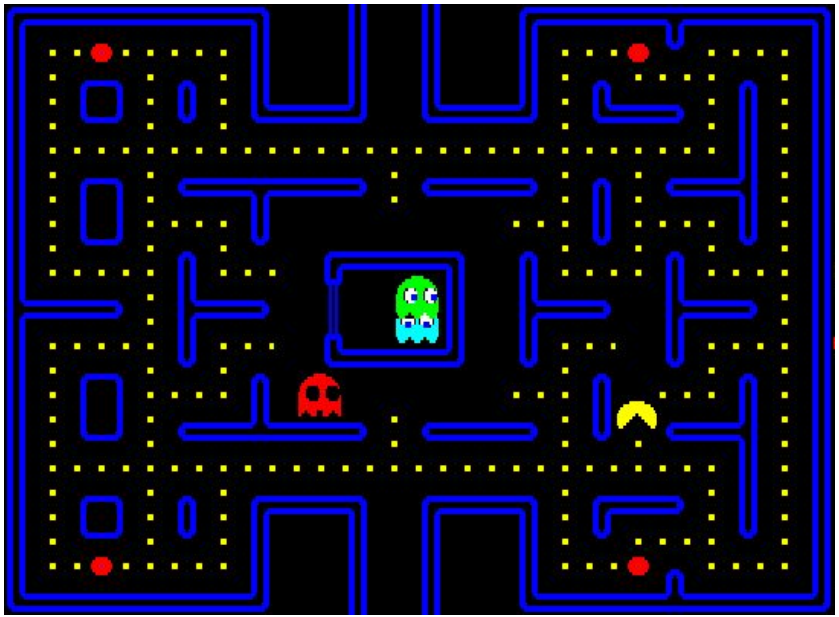
Scraped VGChartz for:

- Global unit sales to date
- Publisher
- Platform/device
- Genre
- Year released

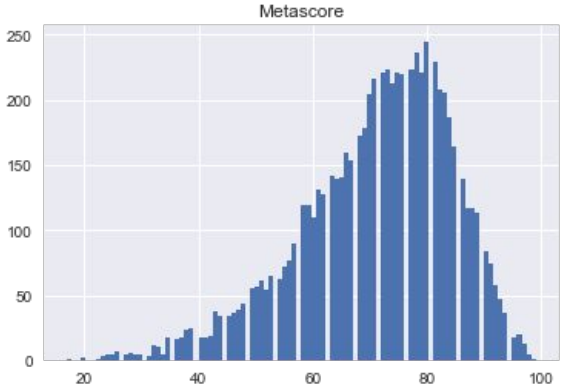
Scraped Metacritic for:

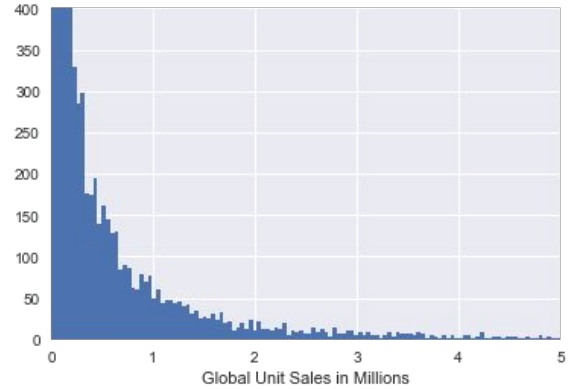
- Average user rating (and # of users)
- Average critic rating (and # of critics)
- ESRB rating
- Number of players

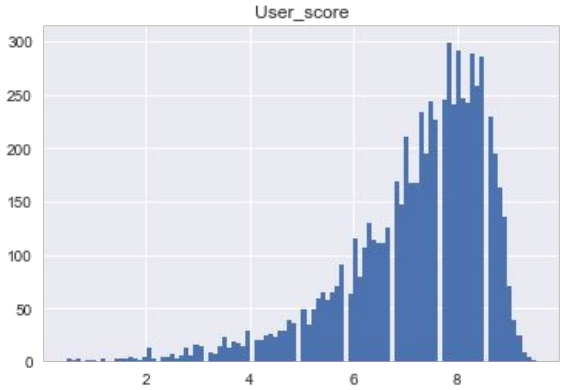
Combined into one dataset and cleaned

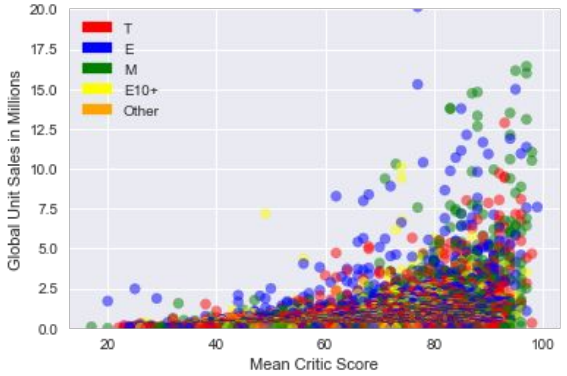


# Exploring the Data




Left skewed critic and user ratings

Critic median = 73/100

User median = 7.5/10

80% of games in dataset had fewer than 1 million in unit sales

Highest selling/rated games mostly rated E or M

# Design


Dropped # of players (too many missing values)

Created dummy variables for:

- Publisher (top 30 and “Other”)
- Device
- Genre
- Rating

Tested simple linear regression model:

- $r^2$ ~.27
- Found it was underfitting

Tested polynomial model to accommodate underfitting:

- Degree 2 & 3
- $r^2$ ~.35 (D2)

# Design cont.

Number of users and critics were dominating models

These factors don't precede sales and shouldn't be used given my business applications. After removing both...

Random Forest to the rescue!

- RF model $r^2$ ~.37
- Gradient boosted model $r^2$ ~.27
- Tweaked with manual and grid searching

Log of Global sales

- RF model $r^2$ ~.53
- Gradient boosted model $r^2$ ~.53

# Results

Random Forest method created the best prediction model

Using log of global sales as the dependent variable created a better (more applicable) prediction model

Most important features:

- Metascore
- User score
- Year
- PC (device)
- Wii
- Nintendo
- Electronic Arts
- Sports
- E (rating)

# Conclusions and Future Research

When Nintendo makes a Wii sports game for everyone, they are going to sell a lot of them...

- Interaction effects suggest certain developers unit sales depend on combinations of the device, genre, and rating
- Critic scores were more important than user scores to predicting sales in all my models
- Next steps
  - Acquire and incorporate number of players, price, sales in first year, user/critic comments
  - Use fuzzy matching to better join data sets
  - Create models for specific developers and devices to see which genres, ratings, etc. perform best for those specific cases


GAME OVER
INSERT COINS
TO CONTINUE