



STEAM®

---

“There just aren’t  
enough people playing  
my wizard simulator”

An analysis of Steam Sales and their impact on  
player numbers

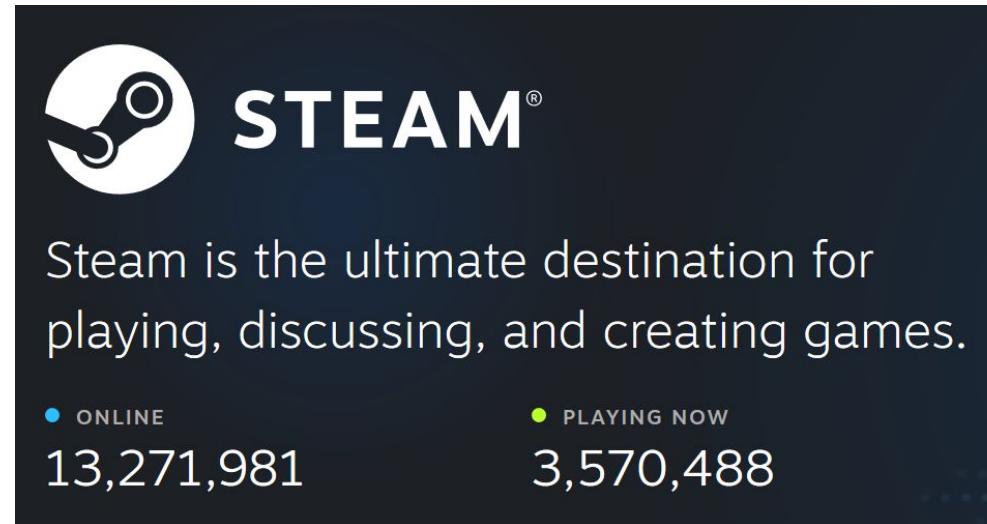
---



# Background & Goals

# Steam is the leading digital distribution platform for PC games

- Over 1 billion accounts
- 47 million daily active users
- Over 25,000 games for sale
- Easily accessible to small developers



# At least twice a year, Steam has major sale events



- Every Summer and Winter
- Around 75% of games go on discount
- Various activities on the storefront
- Gamification of spending money
- Basically a capitalist's wet dream

In 2018, >9000 games were released on Steam,  
an average of 25 games per day

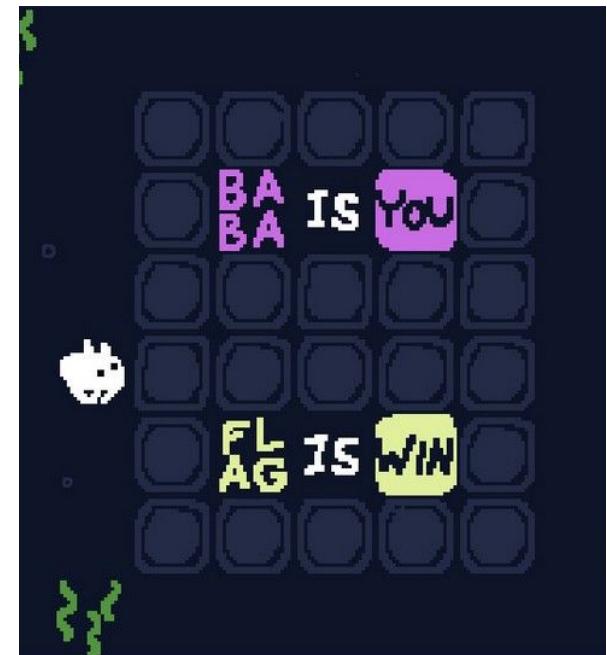


# Goal

Model the impact the Summer Sale 2019 had on player numbers

Impact measured by change in average numbers between June and July 2019

Or at least be a pilot to prove that deeper investigation is worthwhile



A cartoon-style illustration of a garage filled to the brim with various items. In the foreground, there's a yellow star-shaped cushion, a soccer ball, a vintage television set, and a small white robot-like character. Behind them is a massive pile of cardboard boxes, some labeled "SHEILD'S". To the left, there's a trash can and a bucket. On the right, shelves are packed with bottles and containers. The garage has a red brick roof and is set against a clear blue sky.

# Data Collection & Cleaning

# Steam Store API

Provided features for all games on the store

(though actually ended up finding a pre-cleaned version on Kaggle..!)

```
▼ developers:
  0: "CAPCOM Co., Ltd."
▼ publishers:
  0: "CAPCOM Co., Ltd."
▼ price_overview:
  currency: "GBP"
  initial: 4999
  final: 4999
  discount_percent: 0
  initial_formatted: ""
  final_formatted: "£49.99"
► packages: [...]
► package_groups: [...]
▼ platforms:
  windows: true
  mac: false
  linux: false
▼ metacritic:
  score: 88
  url: "https://www.metacritic.com/game/pc/monster-hunter-world?ftag=MCD-06-10aaa1f"
▼ categories:
  ▼ 0:
    id: 2
    description: "Single-player"
  ▼ 1:
    id: 1
    description: "Multi-player"
  ▼ 2:
    id: 9
    description: "Co-op"
  ▼ 3:
    id: 22
    description: "Steam Achievements"
  ▼ 4:
    id: 29
    description: "Steam Trading Cards"
```

## Lovers in a Dangerous Spacetime

[Info](#)   [History](#)   [Stats](#)   [Regions](#)

Released on 09 Sep 2015

821st most popular



Historical Low      Steam on 2018-11-21    67% off    £3.62  
 Current Best      Humble Store on 2019-08-15    0% off    £10.99

Wait for better price  
 Add to Collection  
 Hide from lists (ignore)

Lovers in a Dangerous Spacetime is a 2-player co-op micro-platformer set inside a neon spaceship locked in battle with hordes of space baddies. Players work together running back and forth between ship control rooms, manning turrets, lasers, shields and thrusters to rack up points and stave off a vacuumy demise. In mish-mash terms you could describe it as Jumpman meets

## Log

|                  |          |              |          |        |        |         |     |        |         |
|------------------|----------|--------------|----------|--------|--------|---------|-----|--------|---------|
| 2019-08-15 19:08 | 25 days  | Humble Store | regular: | £10.99 | -      | actual: | 0%  | £10.99 | +150.3% |
| 2019-08-08 19:26 | 6 days   | Humble Store | regular: | £10.99 | -      | actual: | 60% | £4.39  | -60.1%  |
| 2019-08-05 19:16 | 2 days   | Humble Store | regular: | £10.99 | -      | actual: | 0%  | £10.99 | +150.3% |
| 2019-08-01 19:55 | 3 days   | Humble Store | regular: | £10.99 | -      | actual: | 60% | £4.39  | -60.1%  |
| 2019-07-22 20:33 | 49 days  | Steam        | regular: | £10.99 | -      | actual: | 0%  | £10.99 | +150.3% |
| 2019-07-18 20:12 | 3 days   | Steam        | regular: | £10.99 | -      | actual: | 60% | £4.39  | -60.1%  |
| 2019-07-09 20:25 | 8 days   | Steam        | regular: | £10.99 | -      | actual: | 0%  | £10.99 | +150.3% |
| 2019-07-09 00:15 | 2 months | GOG          | regular: | £11.79 | -      | actual: | 0%  | £11.79 | +146.1% |
| 2019-07-08 19:07 | 23 days  | Humble Store | regular: | £10.99 | -      | actual: | 0%  | £10.99 | +150.3% |
| 2019-07-01 19:25 | 6 days   | Humble Store | regular: | £10.99 | -      | actual: | 60% | £4.39  | -60.1%  |
| 2019-07-01 15:00 | 2 days   | GOG          | regular: | £11.79 | -10.0% | actual: | 50% | £4.79  | -50%    |
| 2019-06-25 19:44 | 13 days  | Steam        | regular: | £10.99 | -      | actual: | 60% | £4.39  | -60.1%  |
| 2019-02-18 19:19 | 4 months | Humble Store | regular: | £10.99 | -      | actual: | 0%  | £10.99 | +150.3% |
| 2019-02-14 19:07 | 3 days   | Humble Store | regular: | £10.99 | -      | actual: | 60% | £4.39  | -60.1%  |
| 2019-02-11 19:26 | 4 months | Steam        | regular: | £10.99 | -      | actual: | 0%  | £10.99 | +150.3% |
| 2019-02-04 19:11 | 6 days   | Steam        | regular: | £10.99 | -      | actual: | 60% | £4.39  | -60.1%  |
| 2019-01-28 19:12 | 6 days   | Steam        | regular: | £10.99 | -      | actual: | 0%  | £10.99 | +150.3% |

## IsThereAnyDeal.com

Scraped information on whether games were discounted in the Summer Sale and by how much

|                  |         |       |         |     |       |
|------------------|---------|-------|---------|-----|-------|
| 2019-06-25 19:44 | 13 days | Steam | actual: | 60% | £4.39 |
|------------------|---------|-------|---------|-----|-------|

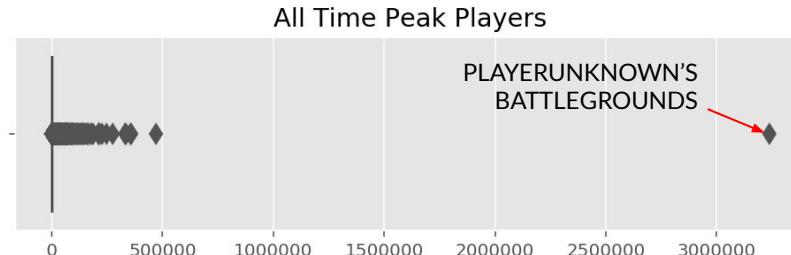
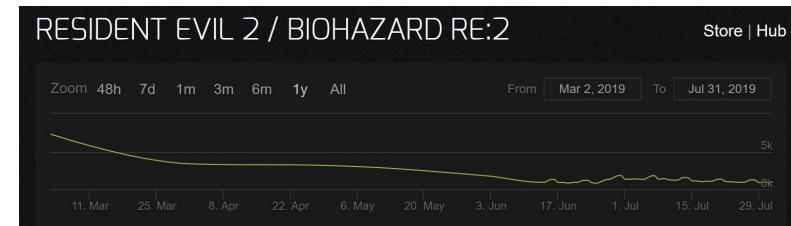
# SteamCharts.com

Scraped information on player numbers by month



# Datasets were combined and cleaned

- Only games released before 2019
  - hopefully ensures “stable” average monthly players
- Cleaned some outliers
  - e.g. from all-time peak players:



# Final dataset included 13,558 games

| index | appid | name    | release_date          | english    | developer | publisher     | required_age  | achievements | total_ratings | ...    | all_time_peak | may_19_av | m       |
|-------|-------|---------|-----------------------|------------|-----------|---------------|---------------|--------------|---------------|--------|---------------|-----------|---------|
| 0     | 0     | 10      | Counter-Strike        | 2000-11-01 | 1         | Valve         | Valve         | 0            | 0             | 127873 | ...           | 65188.0   | 9246.58 |
| 1     | 1     | 20      | Team Fortress Classic | 1999-04-01 | 1         | Valve         | Valve         | 0            | 0             | 3951   | ...           | 191.0     | 58.06   |
| ...   | ...   | ...     | ...                   | ...        | ...       | ...           | ...           | ...          | ...           | ...    | ...           | ...       | ...     |
| 13556 | 22577 | 1001880 | aMAZE Valentine       | 2018-12-28 | 1         | Blender Games | Blender Games | 0            | 1             | 15     | ...           | 2.0       | 0.04    |
| 13557 | 22578 | 1002490 | Roulette Simulator 2  | 2018-12-31 | 1         | JDRumble      | JDRumble      | 0            | 1             | 4      | ...           | 2.0       | 0.02    |

Almost three-quarters of which were discounted:

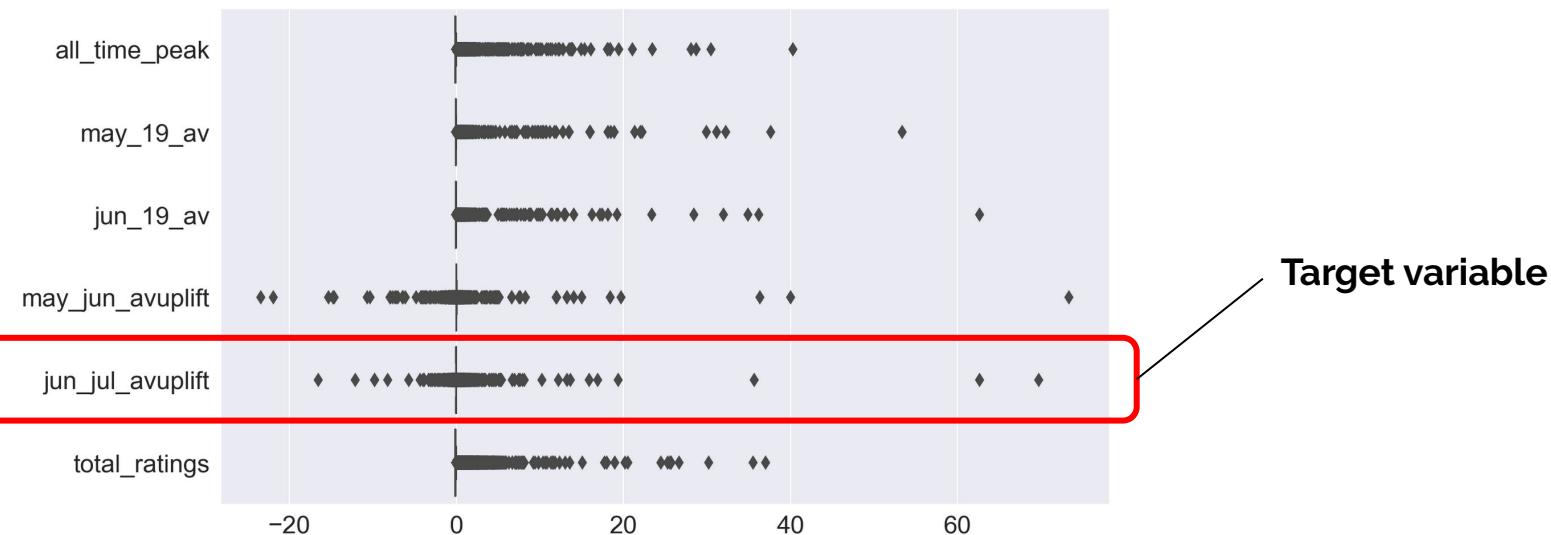
72.8%

27.2%

Discount

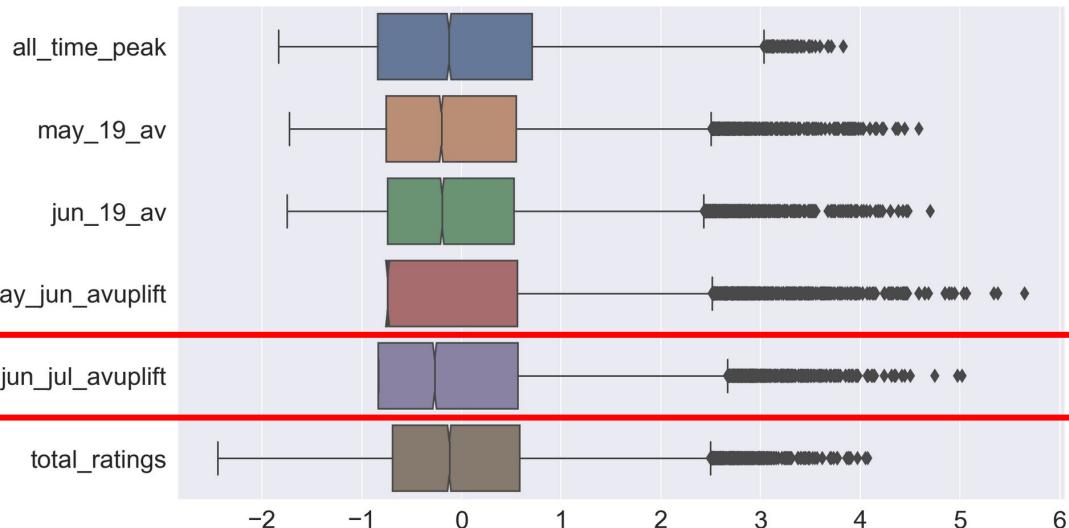
No discount

# EDA showed a challenging distribution of data



Note: data has been scaled to allow for easy comparison

# Logging proved the most successful transformation, though still not ideal



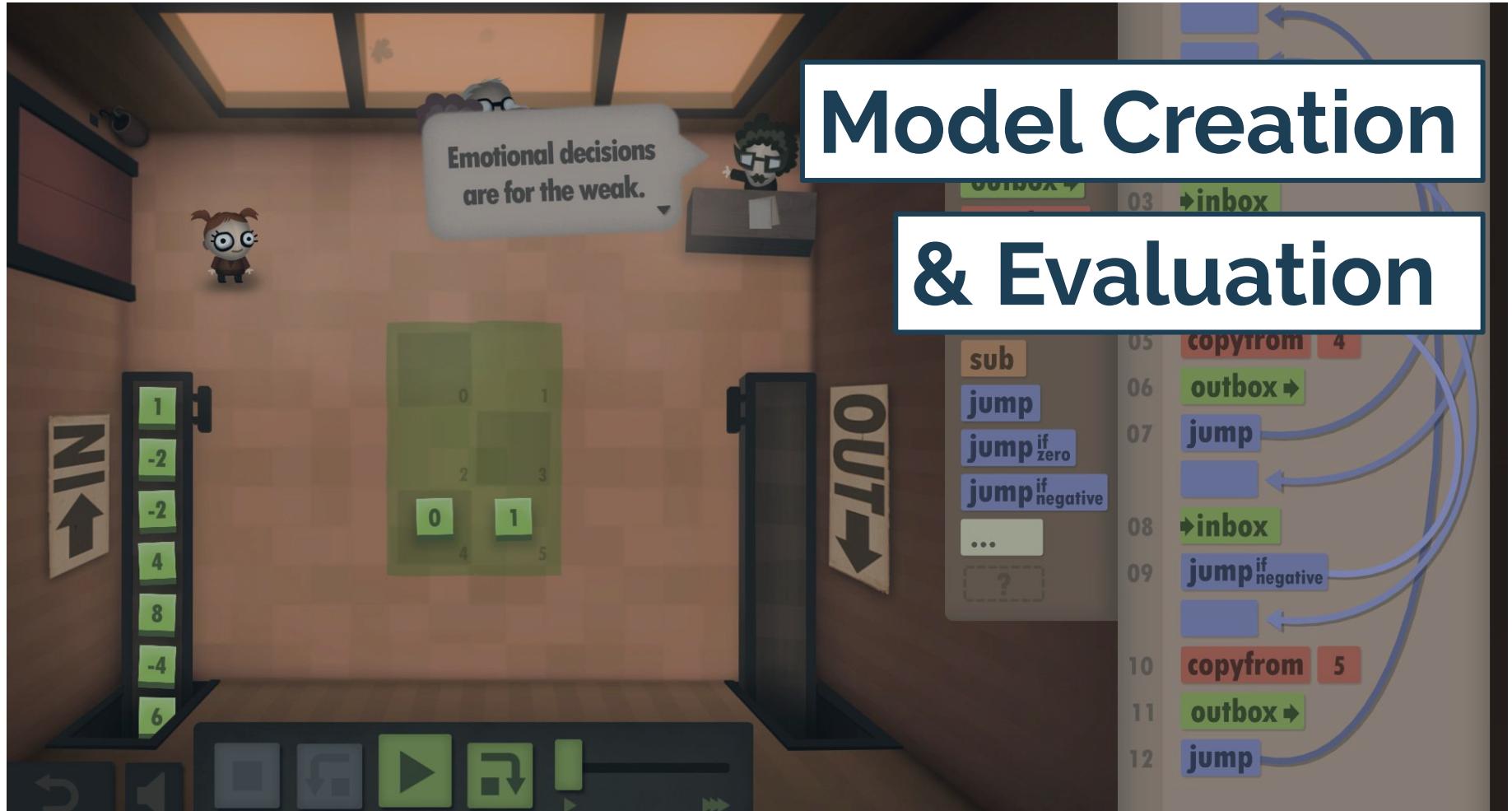
Negative numbers had to be converted to 0s and then have a very small number added

A drop in player numbers is now considered “no increase”

Note: data has been scaled to allow for easy comparison

# Model Creation

## & Evaluation



# Random Forest performed the best of initial models, but still not very successful

R2 Scores:

CV score: 0.350

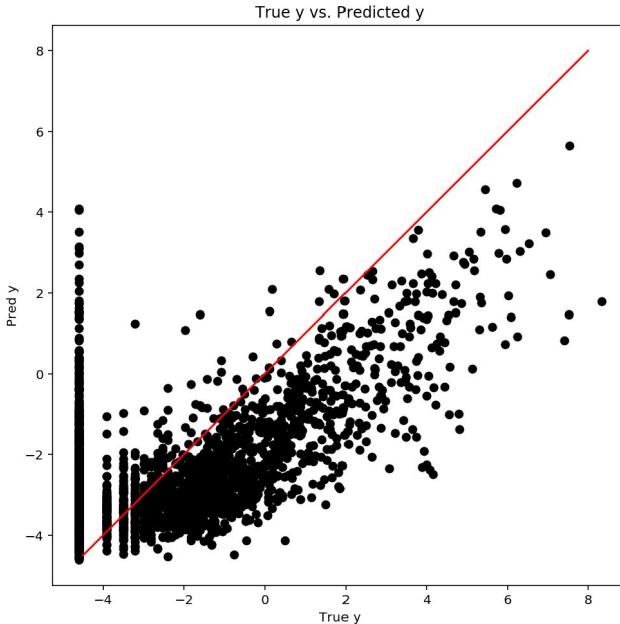
CV std = 0.034

Test score: 0.375

And even worse when the log transformation is undone:

Test score: 0.041

# Digging into the diagnostics, the model was struggling with the “zeros”

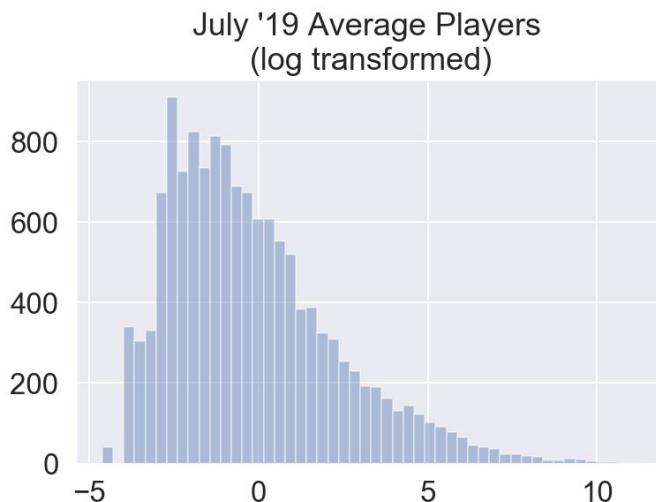


Most of the positive changes are being *underestimated*, while most of the zeros are *overestimated*

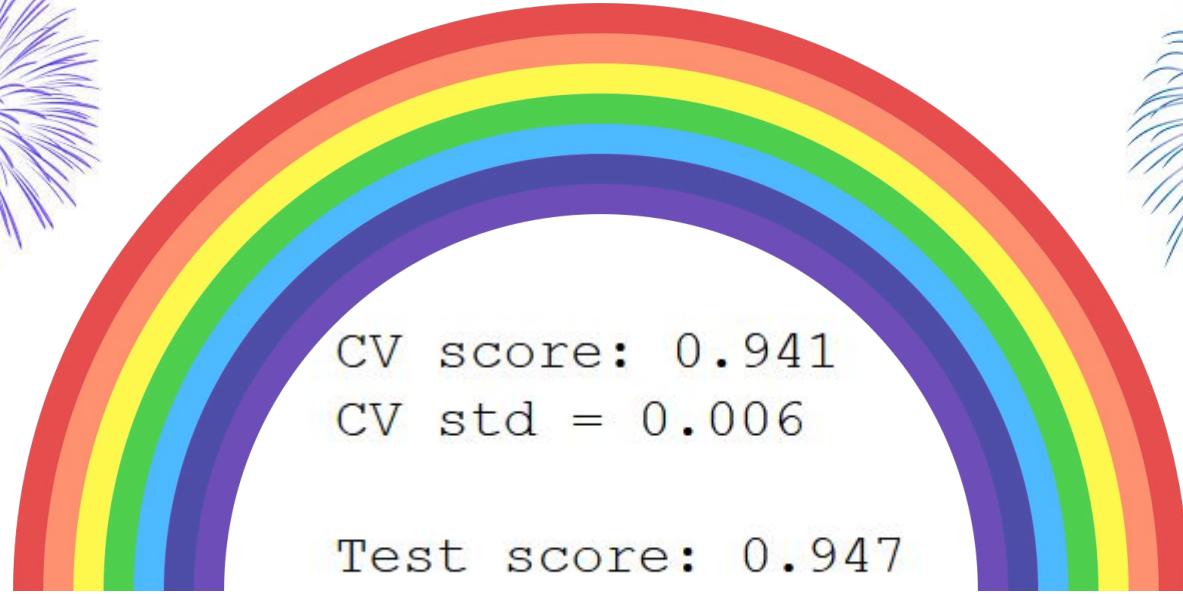
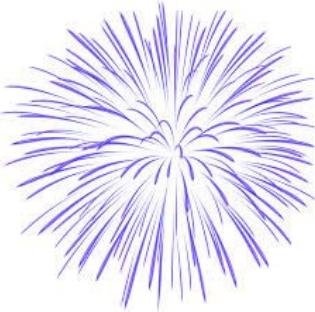
Clearly a new approach was needed...

# The target variable was changed to be the number of players in July

No negative numbers to deal with resulting in a skewed but not awful distribution:



and the test scores....



CV score: 0.941

CV std = 0.006

Test score: 0.947



# But looking at the feature importance, it's clear the score isn't giving the full picture

|               |        |
|---------------|--------|
| jun_19_av     | 0.9323 |
| may_19_av     | 0.0133 |
| all_time_peak | 0.0108 |
| ratingpc      | 0.0064 |
| total_ratings | 0.0061 |
| jun_19_peak   | 0.0060 |
| price         | 0.0049 |
| may_19_peak   | 0.0046 |
| discount      | 0.0032 |
| gen_Casual    | 0.0008 |

Almost all the predictive power is coming from one variable:

The previous month's player numbers

# Instead, the model is compared to a new baseline: the previous month's numbers

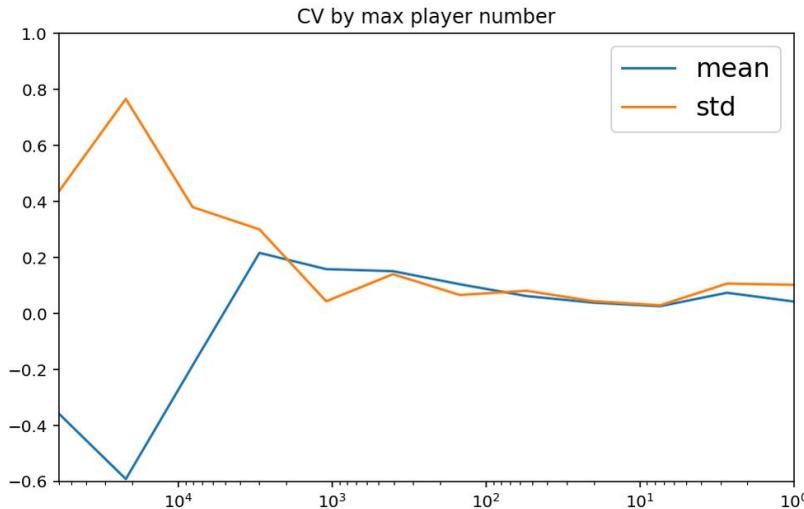
|                                 |      |
|---------------------------------|------|
| Modified R2:                    | 0.22 |
| Modified R2:<br>(actual values) | 0.03 |

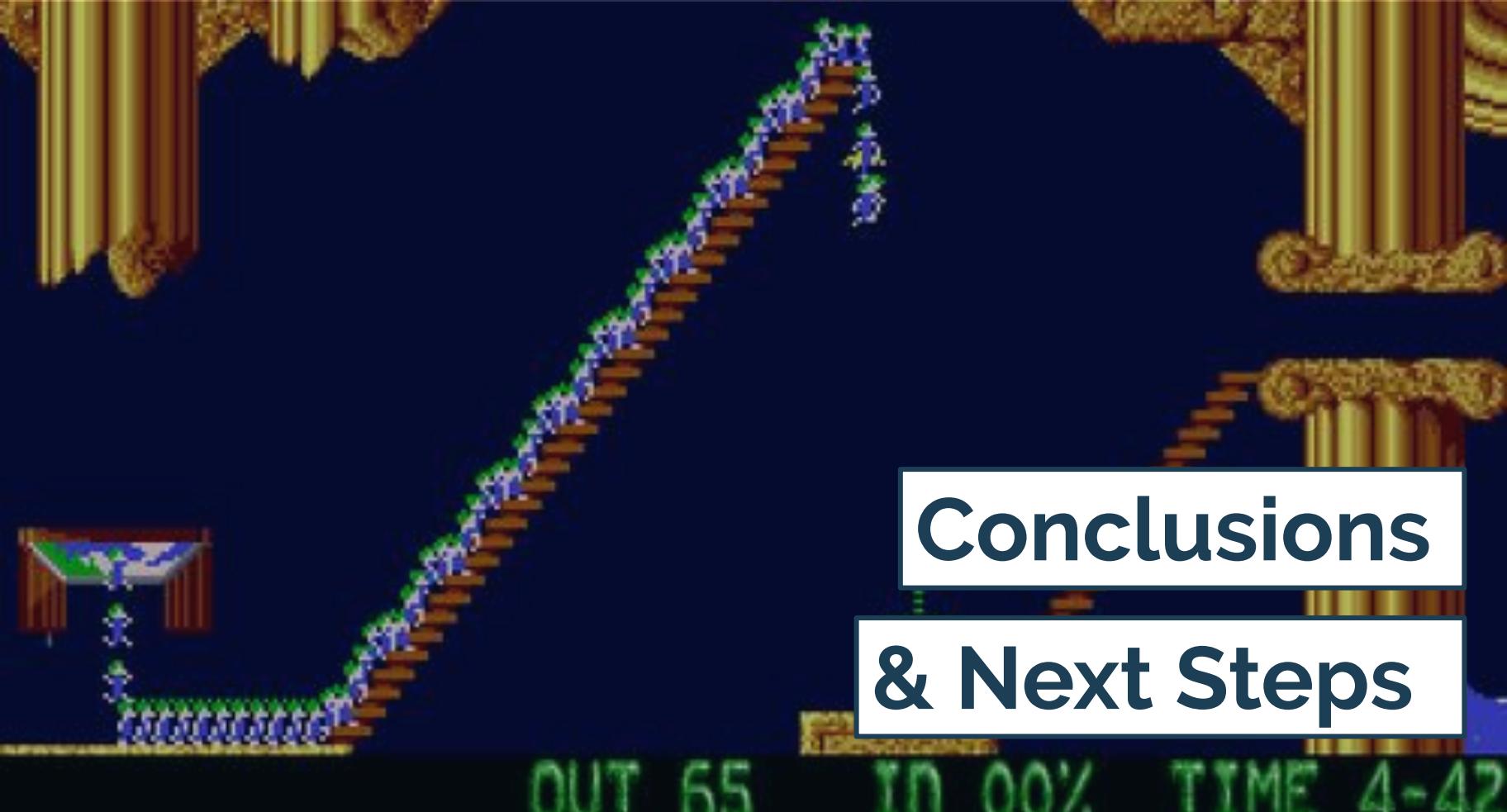
And looking at the residuals, a lot of the error is coming from a few of the most popular games:

## Residuals:

|      | appid  | name                     | discount | jun_19_av | absolute_resid |
|------|--------|--------------------------|----------|-----------|----------------|
| 3272 | 322330 | Don't Starve Together    | 60.0     | 11238.05  | 6125.890418    |
| 649  | 39210  | FINAL FANTASY XIV Online | 0.0      | 14339.24  | 5075.722429    |
| 2308 | 281990 | Stellaris                | 75.0     | 15030.93  | 4188.414862    |
| 17   | 550    | Left 4 Dead 2            | 80.0     | 10076.32  | 2675.761623    |
| 107  | 4000   | Garry's Mod              | 75.0     | 29240.97  | 1809.578003    |
| 1573 | 242760 | The Forest               | 0.0      | 4353.57   | 1497.303871    |
| 1198 | 211820 | Starbound                | 35.0     | 4177.90   | 1282.701823    |
| 6914 | 480490 | Prey                     | 0.0      | 570.80    | 1273.516436    |
| 4465 | 367520 | Hollow Knight            | 40.0     | 2160.58   | 1127.545188    |
| 3006 | 311690 | Enter the Gungeon        | 50.0     | 2699.46   | 1036.924353    |

# Restricting the samples by popularity doesn't really improve accuracy, but does help stability





# Conclusions & Next Steps

OUT 65 IN 00% TIME 4-42

# Predicting change in player numbers is not easy!

- Challenging distribution of data
- Difficult to assess accuracy





Based on the model(s)  
discount has little  
impact on player  
numbers

- Current popularity is the biggest predictor
- Low accuracy, so not a definitive statement
- Player numbers != sales!

# Further investigation is likely worthwhile

- May have more luck collecting data in real-time, not a monthly average
- Can get good accuracy through cheating - maybe a way to use that?
- Just trying new models!



A scene from a video game or animation featuring anthropomorphic cats. In the center, a large white cat with a red bandana and a blue arm guard is shouting and flexing its muscles. It wears a small crown and holds a sword hilt. Behind it, several smaller cats wearing hats and scarves look on. In the foreground, a large crab stands next to several large, rectangular blocks of cheese. The background is a dark, smoky cavern with fire and lava flows.

Thank you!