Metacritic Game Reviews
Exploratory Analysis

Kevin Ramirez

February 2023

DISCLAIMER:

Current analysis is fundamentally flawed due to only partial data gathering (approximately less than fifty-percent scraped). The full Metacritic games list is ordered by Metascore and thus all currently interpreted figures and code results will have a bias towards high scores. Nevertheless the analysis itself should be sound given a complete data set and so I shall continue with examining it as if that was the case.

Introduction

As I have an interest in the different facets of video game culture I originally planned on examining player statistics of those I play often via APIs if they were available. Unfortunately many are either too complex or limited due to player privacy concerns that I did not see it reasonable to follow through in the given time allowance. Thus I decided to explore a few review aggregators and search for any corresponding databases. In my brainstorming I planned to explore these sites with a focus on individual critic scores as well as genres and their potential relationship with other general game statistics. My first pick of these aggregation sites was OpenCritic, however it lacked any genre data whatsoever and I would find it unsatisfactory to have to reference multiple sites for my analysis and potentially reconcile differences. I then noticed that Metacritic did contain this information, yet many preexisting data sets of this sites were either outdated or unsuitable for my use. In particular many did not collect individual critic review scores and only stored the aggregate meta and userscore for each title. It is then that I made the decision to create my own custom data set [Ram23] which proved to be more time consuming then expected. Ultimately the data set that I will be examining in this paper is consistent of three main sections (general statistics, genres, critics). General encompasses game title, platform, release date, metascore, and userscore. The genres column contains sub-columns corresponding to approximately the top fifty most common genres in multi-binary encoding format. Similarly the critics column contains sparse sub-columns corresponding to each publication and their individual scores given for each game.

Inspiration

While this particular data set is unique to my report, some previous research proved to be quite helpful in solidifying what direction to take with the exploratory analysis. In particular two users on Kaggle, Joshua Isanan [Isa20], who worked to predict the number of reviews a given game might have based on various features, and BrunoVR [Rib22] whom focused primarily on platform related exploratory analysis. Bruno is actually the originator of that particular data set [Rib21] which holds many similarities to my own , though I did not end up referencing it very much due to differences in programming methodology.

Analysis

Genre Metascores over Time

To start I was curious as to whether trends and upticks in certain genres on large scales could be notably seen solely through metascore. And while it'd be easily to fall into the temptation of attempting to explain all fifty plus genres, I observed two very extreme examples. The first is one that I can explain with near certainty, the MMO.

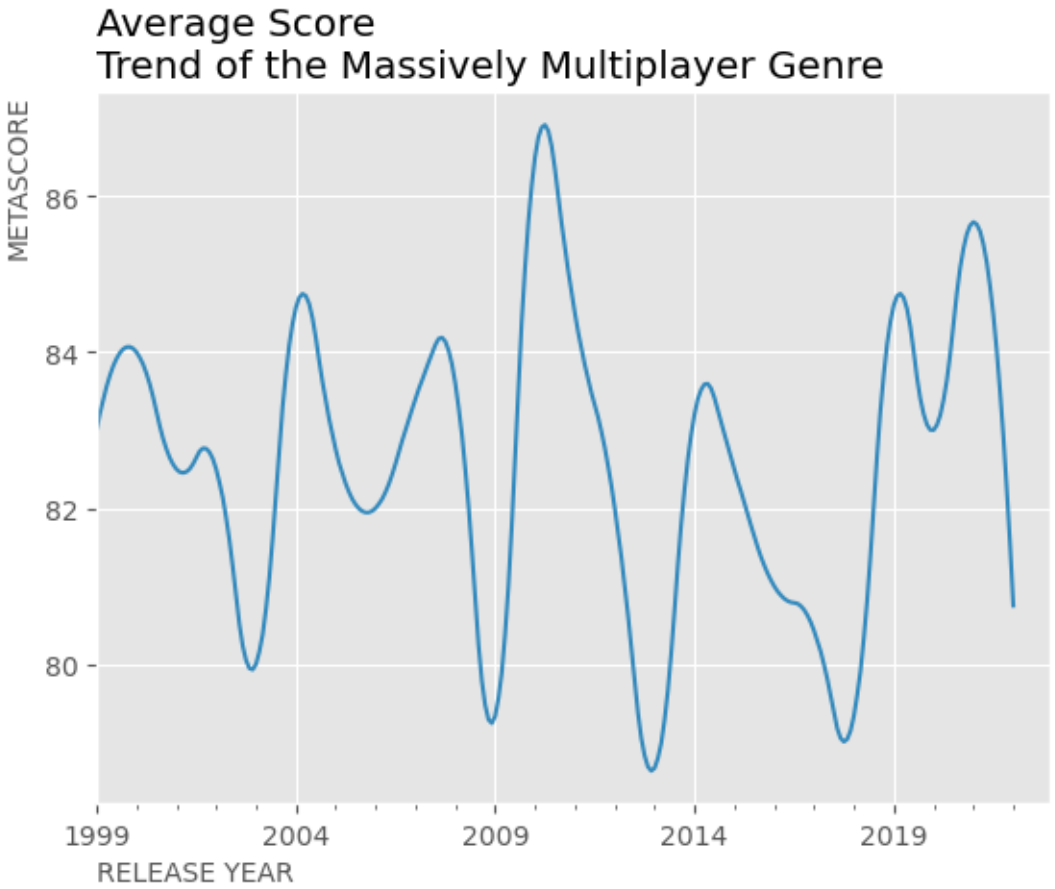


Fig 1. The MMO genre displays a high degree of volatility coinciding with the ebb and flow nature of live service/expansion based games.

As displayed above, the genre is notorious for player lulls while new expansions or events are in development and no new content exists for players to consume. The most prominent peaks of Fig 1. correspond as follows:

- 2004 - The original release of World of Warcraft.
- 2010 - The release of the much acclaimed expansion WoW: Cataclysm.
- 2021 - Endwalker, the final expansion to Final Fantasy XIV's near decade first arc.

The second example is one that is a bit vaguer in nature, though there are several aspects that could play into explaining it.

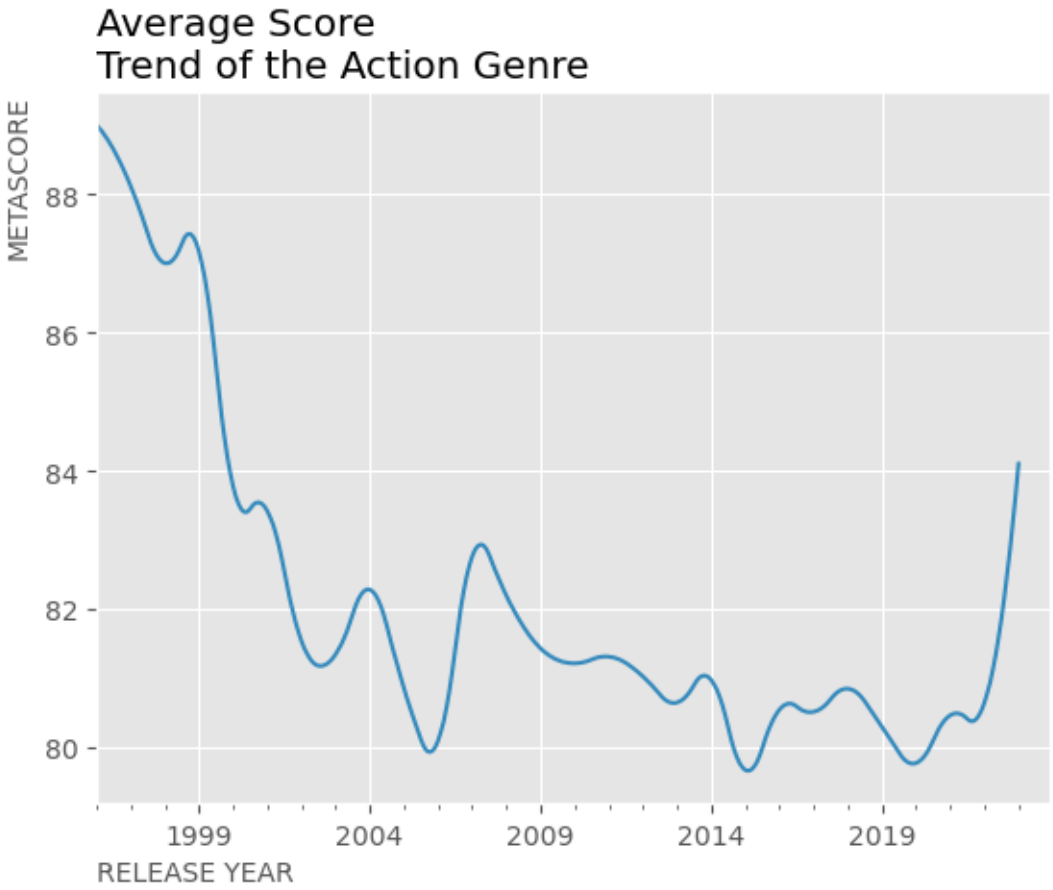


Fig 2. The Action genre has seen a general decline likely due to a loss of novelty as the medium and technology advanced.

The action genre has seen a notable decline which I hypothesize might be due to the genre it self no longer being as revolutionary as it once was with examples like Half-Life (1998) or the original DOOM series often defining the fundamentals of the genre itself. Another explanation is the slow introduction and rejection of Triple AAA style games, which while often engaging and high budget, are often remarked to be formulaic. The later uptick could then be due to a recent return to form within the genre.

It should be noted that the smoothing is purely for graph interpretability and could be disable within the analysis code. Other genres could also be visualized if you so choose.

Critic Scores across Region

Another more niche curiosity I had was whether certain larger publications with internal subdivisions displayed the same general evaluation of games or if they held similarity in name only. Let us first observe IGN one of the most prominent review sites in the industry.

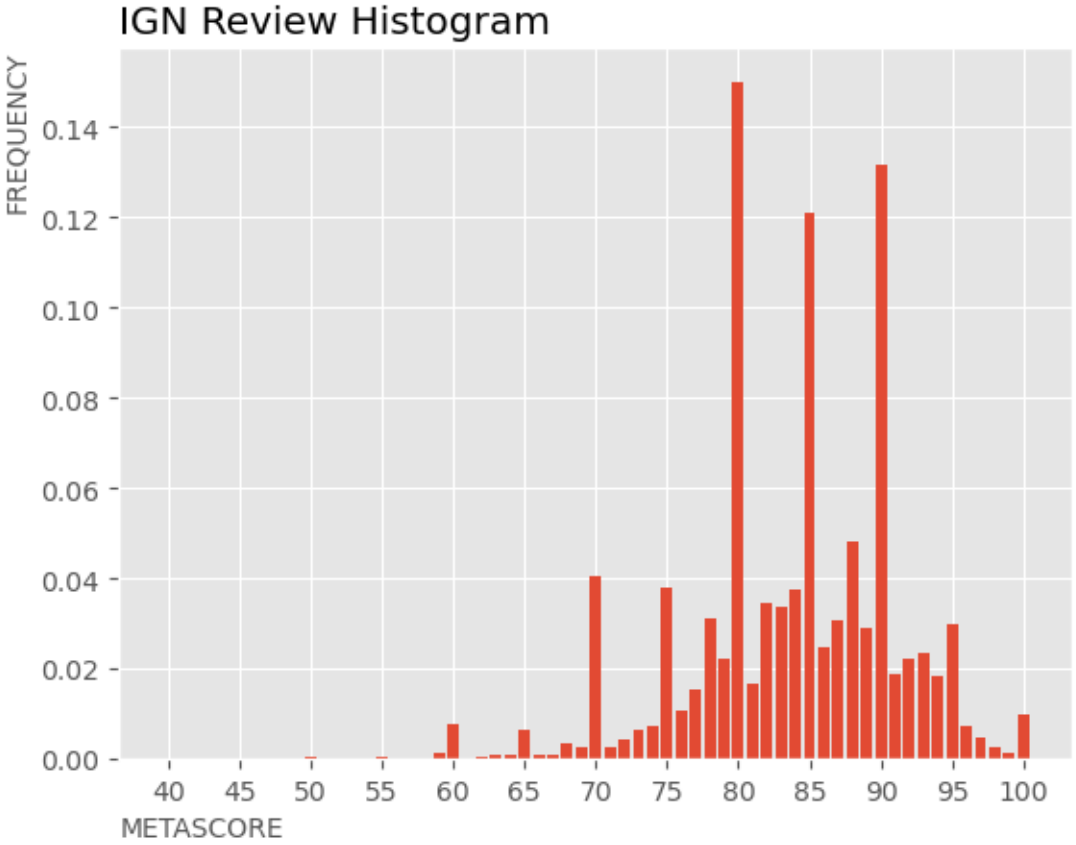


Fig 3. Unsurprisingly, IGN's scoring is fairly unimodal and negatively skewed bringing two notable insights. The first is the mean being fairly offset from a score of exactly fifty, but this is a known systematic review bias (possibly originating from schooling system grades setting an expectation that anything below sixty is failing and not mediocre). The second is the tendency for reviews to fall on values divisible by five. This is a widely noted aesthetic quirk backed by psychology research.

Next we can compare the array of scores given with its subsidiaries.

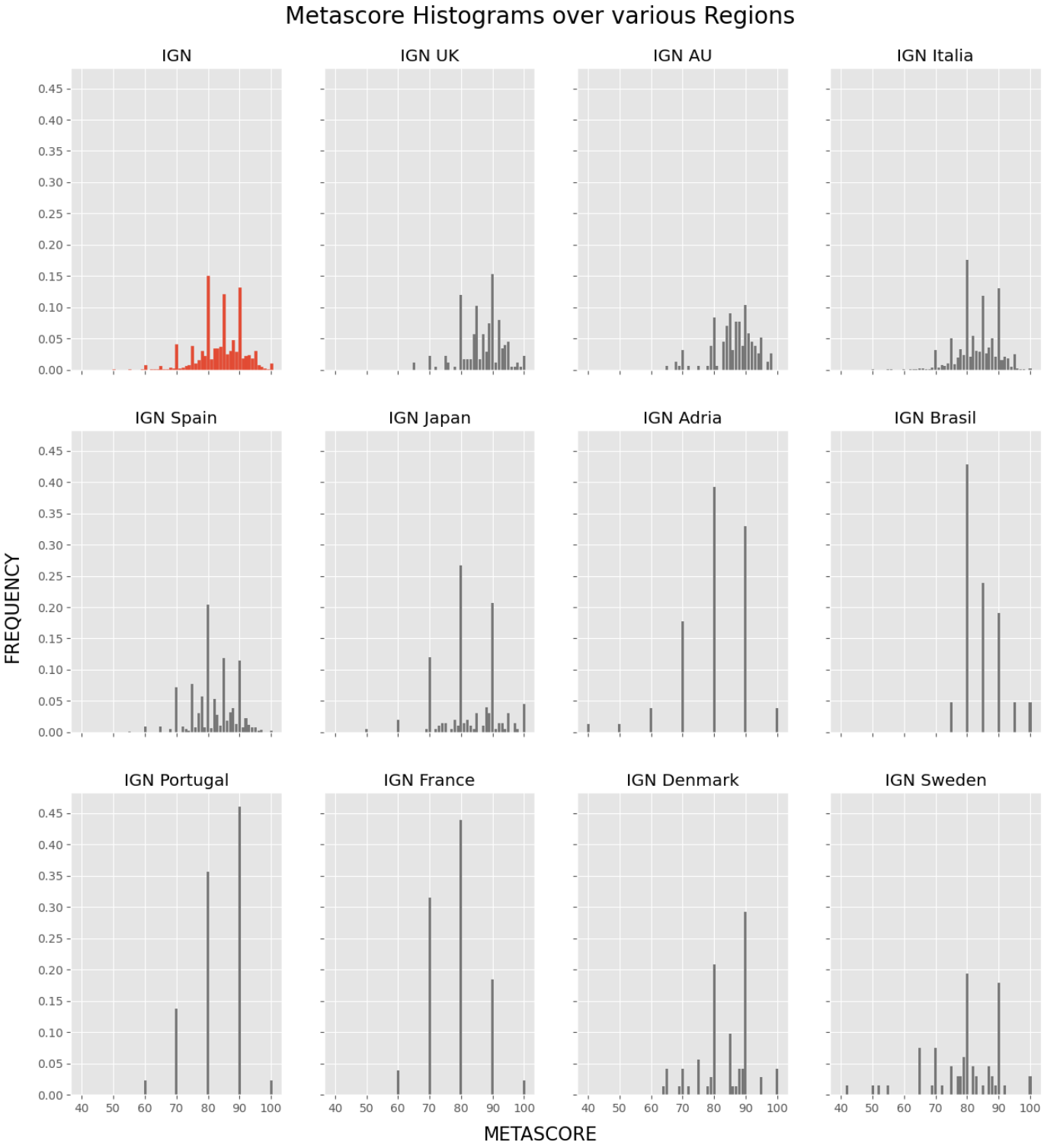


Fig 4. The divisibility/"clean" numbers phenomenon is even more pronounced as other regions seem to be less granular with the ratings they dole out. The smaller regional critics also appear to be fairly consistent with the main publication with the exception of Adria and Sweden which gave lower scorings more often.

Overall they do appear to be in line with one another aside from the fact that the smaller sections are less precise with their scoring.

Critic Scoring Bias

Continuing with a focus on large critic publications, I seeked to compare how reviewers stand against their peers. Which reviewers are more lenient or harsher than expected? As there are over a hundred reviewers I shrunk the scope to the top five most prolific publications while still including the rest via a baseline.

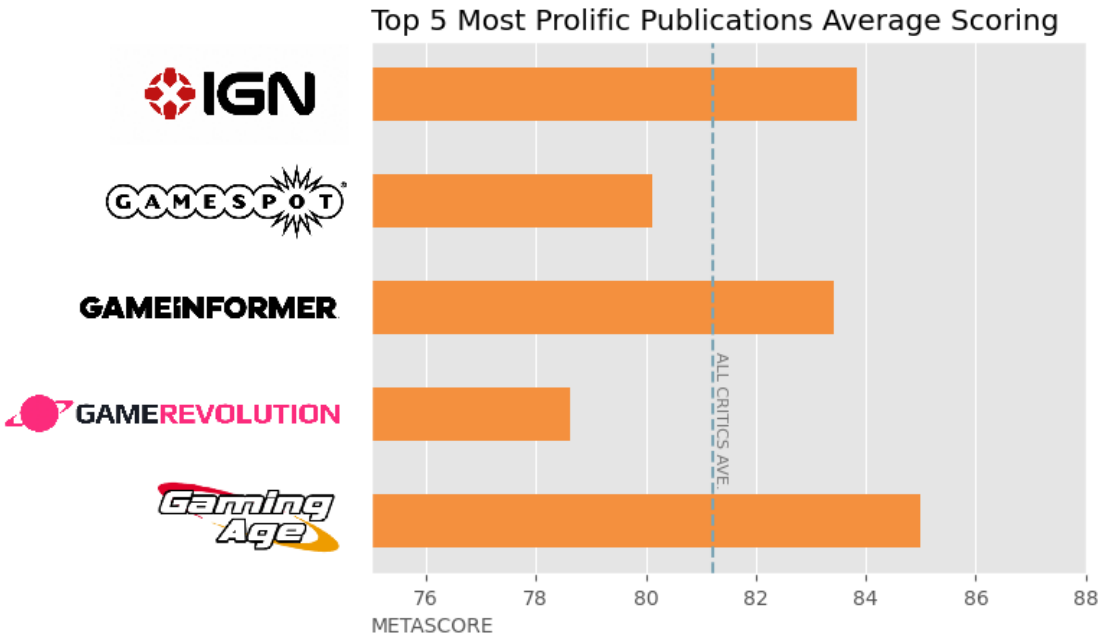


Fig 5. Both Gaming Age and IGN tend to be more lenient in comparison to other critics. Conversely Game Revolution appears to be more critical than the standard.

Game Scores across Genres

Moving on from any one specific reviewer, I decided to observe how critics as a whole rate certain genres in contrast to users. Moreover, the form by which I visualized this relationship also helps us examine any notable differences between the genres themselves. Again I aimed to capture a subset of the full range of genres as to not overload the graph visual space.

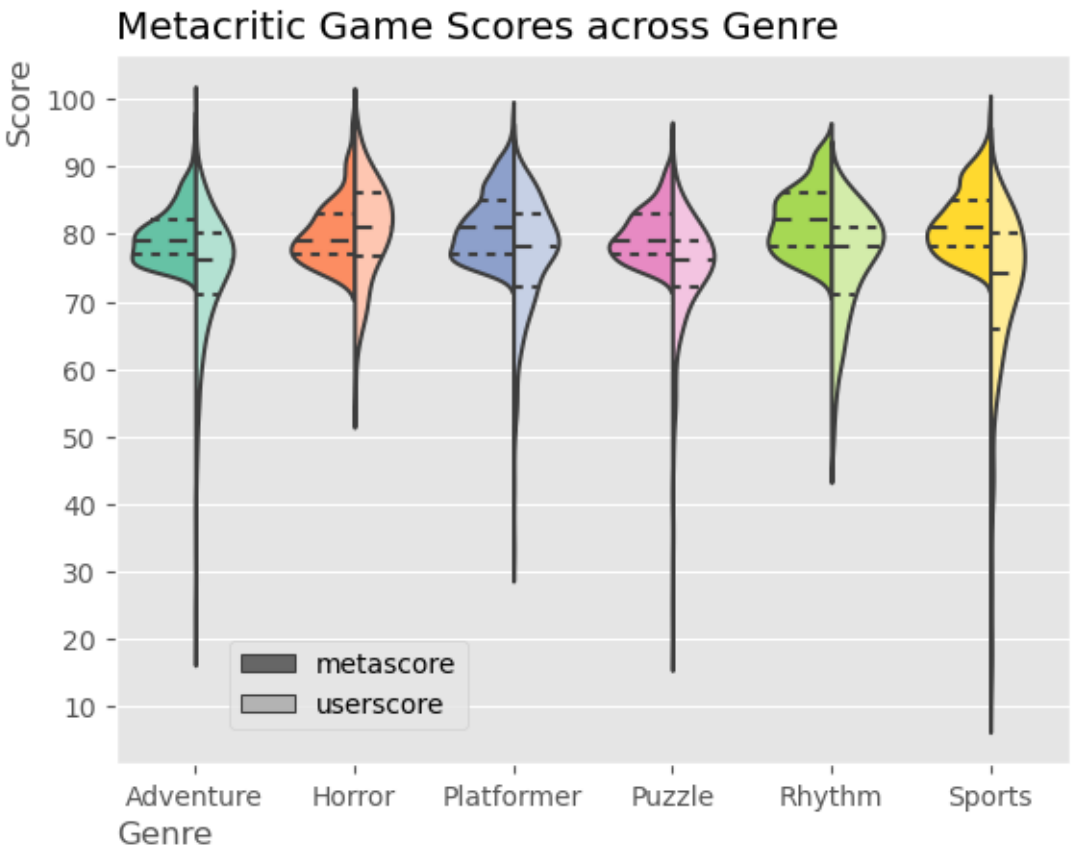


Fig 6. Critics appear to form a much clearer consensus as their interquartile is generally narrower than that of users. They also tend to score higher on most of the genres analysed with the exception of horror. Most interestingly though, is that userscores mostly approximate a heavily skewed unimodal while critic distributions appear more plateau-esque/multimodal-ish.

For example, aside from what is already mentioned in the caption, Horror and Rhythm games tend to be evaluated more favorably by users and have a much less skewed distribution, closer but still not quite as narrow of a band as critic metascores. There also seems to be very little variation overall in terms of any one genre having an outlier median or interquartile range.

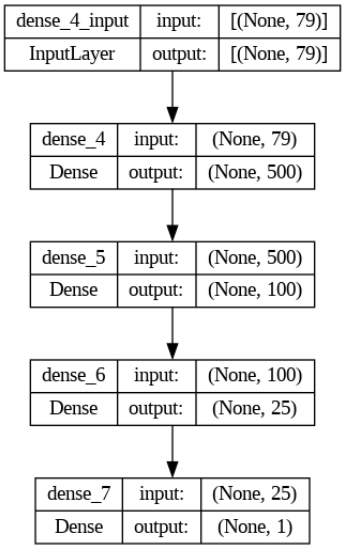
Predicting Critic's Metascore based on Genre and Platform

As seen from the subsection title my final exploratory analysis was an attempt at regression on some of the features of this data set. In particular the data below is attempting to predict the scores from the publication GameSpot. The first method I used was logistic regression which was overall quite subpar.

```
model = LogReg(max_iter=200).fit(train_x, train_y)
print(f"Logistic Accuracy: {model.score(test_x, test_y)}")
✓ 4.3s

Logistic Accuracy: 0.22884615384615384
```

I then moved towards using a dense neural network to more accurately predict the scoring. It's exact structure of (22 OHE platforms and 57 Multilabel encoded genres) features and the subsequent results are depicted in the images below.




```
Epoch 1/20
114/114 [==============================] - 1s 3ms/step - loss: 1714.1760
Epoch 2/20
114/114 [==============================] - 0s 3ms/step - loss: 110.9070
Epoch 3/20
114/114 [==============================] - 0s 2ms/step - loss: 80.6473
Epoch 4/20
114/114 [==============================] - 0s 2ms/step - loss: 67.0019
Epoch 5/20
114/114 [==============================] - 0s 3ms/step - loss: 61.5979
Epoch 15/20
114/114 [==============================] - 0s 2ms/step - loss: 54.1772
Epoch 16/20
114/114 [==============================] - 0s 2ms/step - loss: 53.3625
Epoch 17/20
114/114 [==============================] - 0s 2ms/step - loss: 51.7939
Epoch 18/20
114/114 [==============================] - 0s 2ms/step - loss: 51.5644
Epoch 19/20
114/114 [==============================] - 0s 2ms/step - loss: 51.9122
Epoch 20/20
114/114 [==============================] - 0s 2ms/step - loss: 50.2242
49/49 [==============================] - 0s 1ms/step - loss: 67.6311
```

and tested on a single game:

```
611    It Takes Two
Name: title, dtype: string
1/1 [==============================] - 0s 63ms/step

Predicted GameSpot score: 90.28
Actual score given: 90
```

For further testing and verification I urge the reader to explore the .ipynb file at their own pace.

References

[Isa20] Joshua Isanan. *Predicting the number of Reviews of a Video Game*. Version 11. 2020. URL: `https://www.kaggle.com/code/joshuaisanan/predicting-the-number-of-reviews-of-a-video-game`.

[Ram23] Kevin Ramirez. *metacritic-game-analysis*. Version 1.2. 2023. URL: `https://github.com/kevinramirez723/metacritic-game-analysis`.

[Rib22] Bruno Ribeiro. *Games-Cleaning-and-EDA*. Version 1. 2022. URL: `https://www.kaggle.com/code/brunovr/games-cleaning-and-eda`.

[Rib21] Bruno Ribeiro. *projectGames*. 2021. URL: `https://github.com/BrunoBVR/projectGames`.