Author: Dominion Samuel
CMPT 353 Summer 2025

WHAT MAKES A GAME LOVED?

A MULTI-ANGLE ANALYSIS BACKED BY DATA SCIENCE

INTRODUCTION:

In an industry where millions of games compete for attention, understanding what makes a game resonate with players is vital. In this project, I defined "loved" games as those with consistently high average user ratings. While critic reviews can provide insight, they often vary widely and may not reflect the majority opinion, so this analysis relied primarily on user ratings.

My goal is to explore which features consistently appear in positively rated games, and whether such patterns can be used to predict future success. This report outlines my approach, key findings, and final recommendations. I began this process by cleaning and preparing the data, then proceeded to analyze engagement, timing, maturity, genre, and accessibility, before building predictive models.

DATA AND CLEANING OVERVIEW:

The dataset was extracted from the RAWG API and originally contained data on 10,000 video games. During the cleaning process, duplicate games were removed, and null values were addressed. Games with missing release dates were dropped due to their small proportion, while 'Metacritic' scores were missing in over half of the dataset and retained only for supplementary analysis which is not addressed in this report.

Key features were engineered for analysis. From the release date, year, month, and quarter were extracted. For genres, which were initially stored as lists, the number of genres per game was counted, and each genre was binarized. Platform data was restructured to focus on two dominant types: PC and consoles (PlayStation, Xbox, and Nintendo). A new 'platform type' column was created with four categories—*pc*, *console*, *multi platform*, and *other*. Additional features included the number of platforms and the number of available stores per game.

Games with clearly incomplete metadata, such as zero genres, platforms, or stores were excluded. Titles with projected release dates later than 2024 were also removed due to data incompleteness. Finally, to reduce noise from outliers, games with fewer than 10 user ratings were excluded. The cleaned dataset was saved to CSV for future analysis with around 8000+ games.

ANALYSIS SECTION:

The following sections cover analysis that answer several questions about the dataset using similar but different techniques. All questions hope to provide more understanding as to how user ratings are influenced by several features.

A. Does Engagement Correlate with Love?

Engagement metrics such as *added count, review count, user ratings count, and average playtime(hours)* are recorded after release and are not known during game development. Also *added* only accounts for games added to the RAWG library. Thus, I considered them secondary factors and only related them to the average user rating to find any hidden trends. Their distributions are highly left skewed with most games having low to medium engagement and few outliers.

The scatterplots in Figure 1 use a log scale due to this skew, but average user ratings were kept in their original scale for interpretability. Overall, higher engagement aligns with higher average user ratings, indicating a somewhat positive correlation. However, *average playtime* shows a wider rating variability, suggesting players may spend time on games they do not really love, possibly because of friends, competition or addictive elements. Therefore, engagement, especially hours of playtime, may not fully represent a game's love and reception.

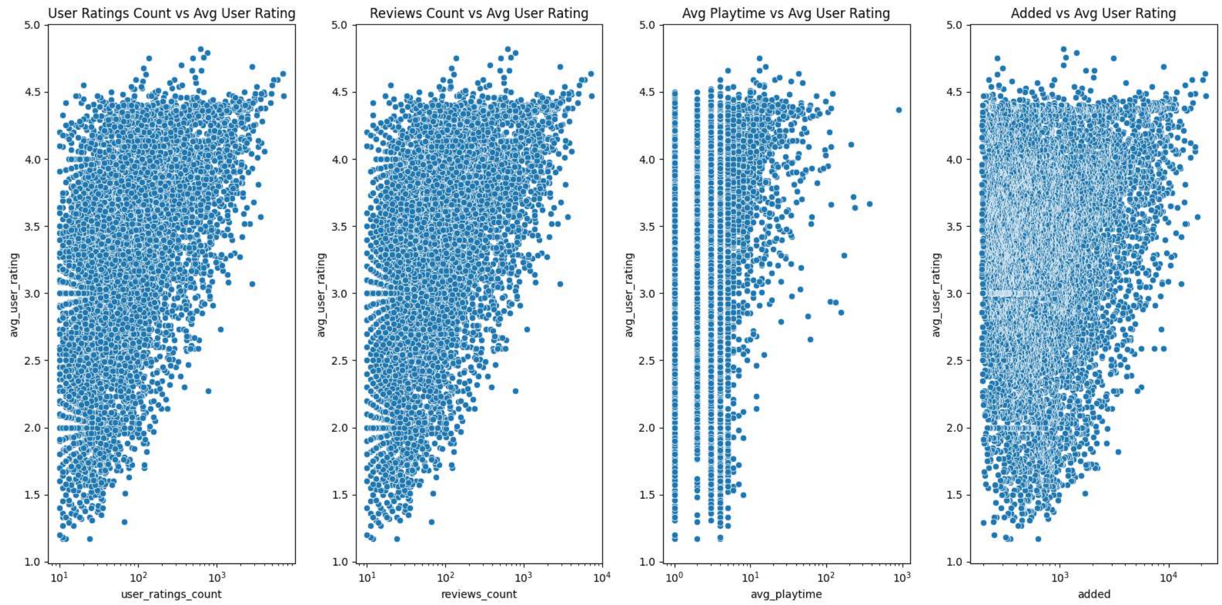

Figure 1: Scatterplot of Engagement Features vs Average User Rating

B. Effect of Release Timing

Release timing was analyzed using extracted release month and quarter. I placed focus on the month and quarter as they are crucial in release timing for game companies. Average user ratings by month in Figure 2 showed relatively stable trends, with games released in April and May having the lowest average ratings. In contrast, games launched in winter months tend to have higher ratings, likely due to holiday season releases benefiting from increased player engagement and marketing efforts. The average rating by quarter supported this with the 4th quarter having the highest average of 3.34 and the 2nd quarter with a low of 3.2. It should be noted that no statistical test was performed to affirm this, so results are only exploratory.

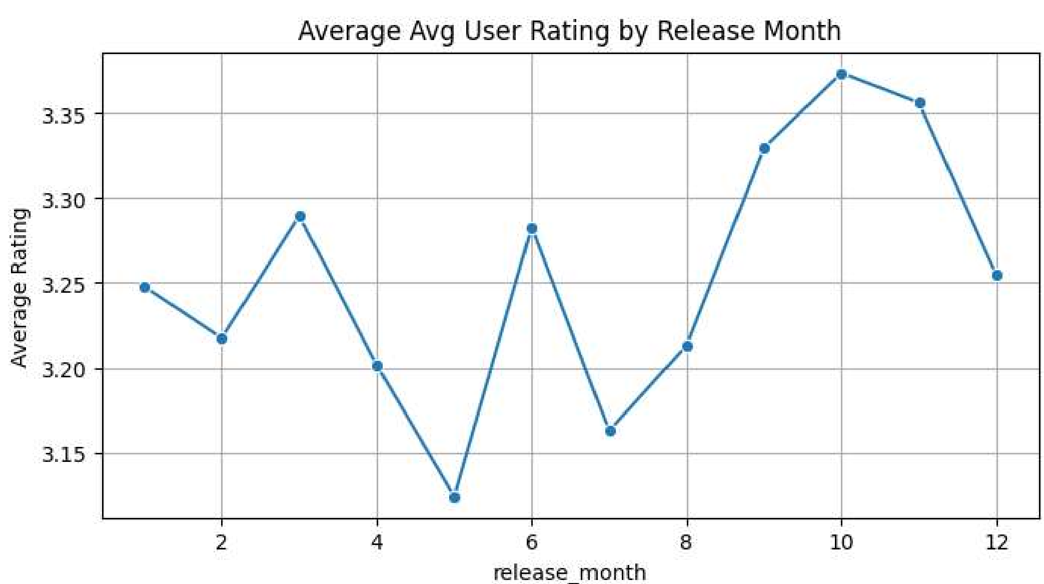

Figure 2: Line plot of Average User Rating by Month

C. Does Maturity Rating Matter?

It is important to note that out of 8,000+ games, only ~3,900 included an ESRB rating. This smaller subset means the findings may not fully represent the entire dataset. Counting the number of games per category revealed that Teen-rated titles were the most common, aligning with their broad target audience. Surprisingly, Mature-rated games ranked second, exceeding Everyone-rated titles. To compare how ratings differ across categories, I focused on Teen, Mature, and Everyone, as these have clearer distinctions without too much overlap in target groups. Boxplots in Figure 3 showed that Mature-rated games had the clear highest median rating, significantly higher than both Teen and Everyone.

Since normality and equal variance assumptions were not met, I used a Kruskal–Wallis test instead of ANOVA to assess whether these differences were statistically significant. The result ($p = 2.27e^{-12}$, $\alpha = 0.05$) allowed rejection of the null hypothesis, indicating that at least one group's median rating differs from the others. Post-hoc pairwise comparisons (Bonferroni-adjusted) revealed that Everyone and Teen ratings do not significantly differ, while Mature games are rated significantly higher than both groups. This supports the visual trend observed in the boxplots, Mature-rated titles consistently achieve the highest ratings in this subset.

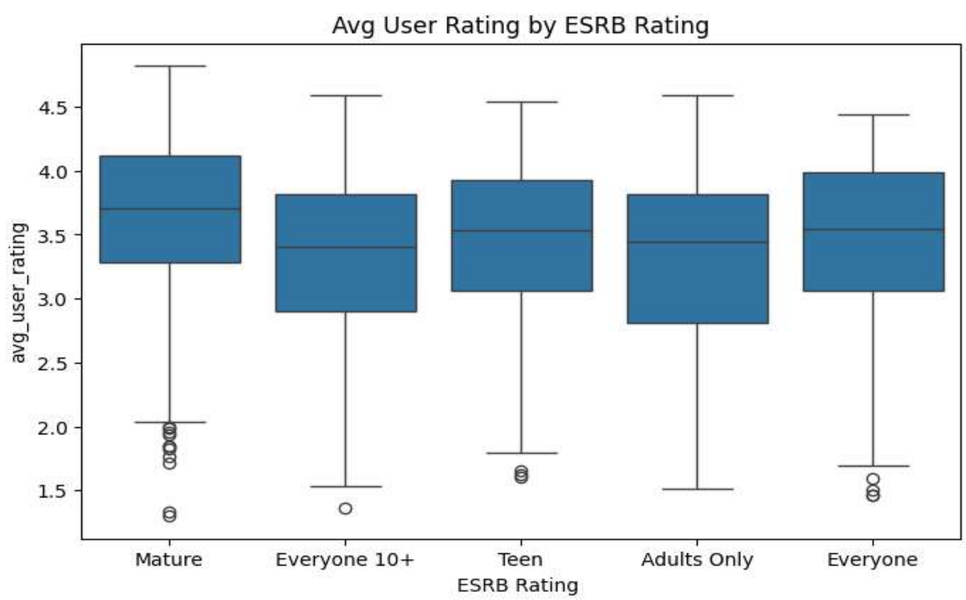

Figure 3: Boxplots of Average User Rating by ESRB Rating

D. Genre Analysis:

To examine trends across game genres, genres were binarized into separate columns and counted for frequency. Action and Indie emerged as the most common genres, while Educational, Card, and Family games appeared far less frequently. For quality comparisons, I calculated the average user rating by genre, restricting the analysis to genres with more than 100 games to avoid inflated scores from underrepresented categories. Surprisingly, Indie games, despite their strong fan following, did not make the top 10 in ratings, and Action games only narrowly secured a spot. Instead, Platformers and Fighting games claimed the top two positions, despite being produced far less often. This suggests that standing out in a niche genre may yield higher player appreciation than following popular trends at least in this dataset.

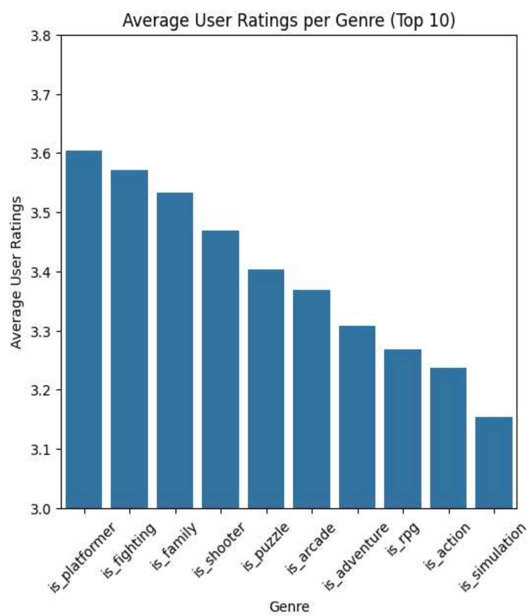

Figure 4: Average User Ratings by Genre

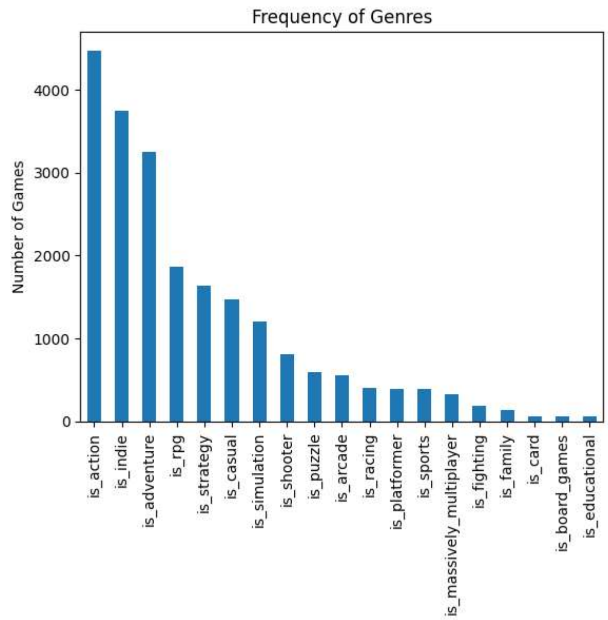

Figure 5: Number of Games by Genre

E. Platform and Store Reach Factor:

For this section, I started with analyzing how reach and availability affect a game's reception. As expected, games available on more stores and platforms tend to be more favored, reflecting gamers' appreciation for ease of access. Focusing on the dominant platform types (PC-only vs. multi-platform, which together made up ~92% of games in this dataset), boxplots revealed that multi-platform games consistently had higher average ratings than PC-only titles. I confirmed this difference with a Mann–Whitney U test (chosen due to non-normal distributions and unequal variances) and got a highly significant value ($p = 6.2e^{-181}$, $\alpha = 0.05$). This shows that games appearing on both consoles and PC tend to receive higher user ratings than games only on PC.

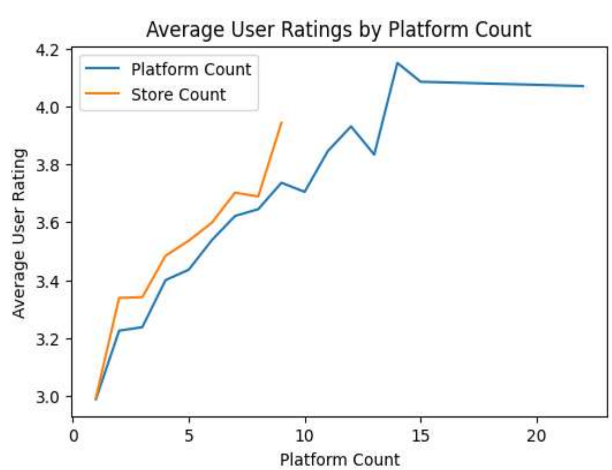

Figure 6: Average User Ratings by Store and Platform Count

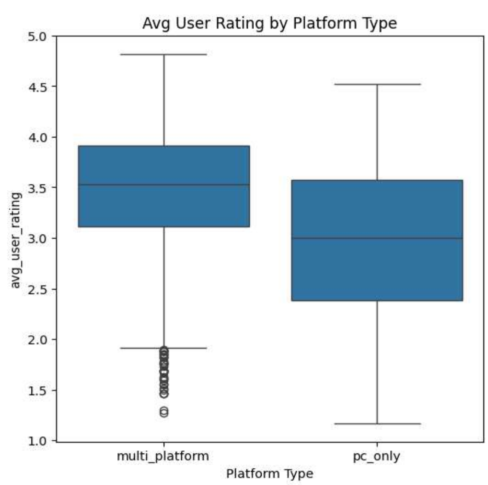

Figure 7: Boxplots of Multi-Platform vs PC Only

MACHINE LEARNING MODELING

REGRESSION: I first tried predicting exact user ratings using a Random Forest Regressor with features available before release (that is excluding engagement metrics). Random Forests were chosen as they can handle model nonlinearity, mixed features and ability to extract feature importances. After standard scaling and one-hot encoding categorical variables like platform, release timing, and maturity, the model's performance remained low (around 28% accuracy), suggesting that user ratings depend on additional factors not captured in this dataset—such as publisher reputation, advertising, and player sentiment from trailers and promotions.

CLASSIFICATION: I then reframed the task as a classification problem, labeling games with average ratings above 4 as "loved." Using the same features and preprocessing, I trained a Random Forest Classifier on a balanced dataset and achieved about 69% accuracy on validation (70% on training). This was significantly better than the regression score and indicates potential for improvement with additional relevant data and features as stated earlier. Beyond accuracy, I inspected the confusion matrix in Figure 8 and examined precision/recall tradeoffs to understand practical utility. The model seems to be having the most issues with incorrectly labelling a game as loved. I also extracted feature importances and inspected the top 15. Accessibility measures (platform/store counts and platform type), maturity rating, and a small set of genre indicators emerged as among the most useful features for the model which supports the earlier findings.

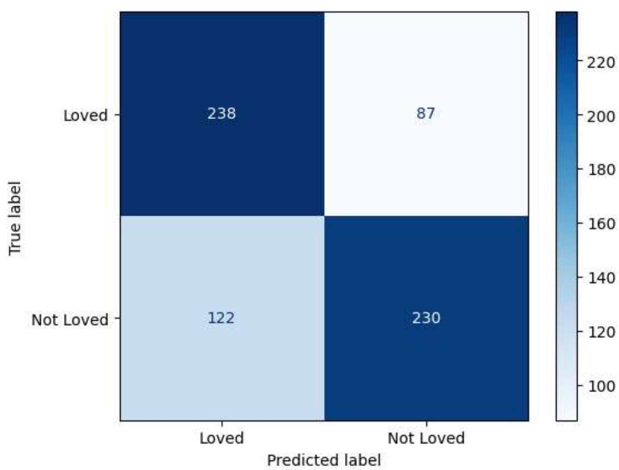

Figure 8: Confusion Matrix for Random Forest Classifier

SUMMARY AND KEY INSIGHTS

The following key insights summarize important trends and findings that emerged from the statistical and modeling approaches applied:

- Games released in the last quarter of the year tend to receive higher ratings and player appreciation
- Mature games have significantly higher ratings compared to other ESRB Categories
- Generally, more engagement correlates with higher ratings, but does not always translate to 'player love' especially when considering playtime
- Genre trends suggest there lie value in creating stand out games in niche or less popular categories rather than flooding the market with more popular genres.
- Accessibility matters to the player base which is shown through positive correlations between the game's reach across platforms and stores and its reception.

LIMITATIONS AND FUTURE WORK

DATA LIMITATIONS: The first was with the report size, I had done much more analysis in my Jupyter Notebooks but had to significantly trim down for the report. I also had a vision for incorporating sales data. However, accessing reliable sales data proved challenging, and attempts to integrate data from sources like '*VGChartz*' significantly reduced the dataset size, limiting analysis.

FUTURE WORK: If I were given more time and report allowance, I would incorporate more extensive critic ratings to explore gaps between critic and user opinions. Additionally, I would investigate deeper into engagement metrics and genre-specific trends. I would also analyse how features like genres, platforms, and yearly trends correlate with each other and how that affects ratings. Lastly, I would love to work more on the classification model. I believe it could get far better than a 69% score if given more useful features. I would also analyze the precision, recall and explore other ways of transforming the features to optimize the model. The primary challenge would be sourcing high-quality data but if found the analysis would prove beneficial.

PROJECT EXPERIENCE SUMMARY:

In this project, I achieved the following:

- Collected and cleaned a raw dataset of 10000+ games from the RAGW API, handling missing values and duplicates in Python with Pandas, resulting in a reliable dataset of 8000+ entries for analysis.
- Engineered pre-release features such as platform type, store count and genre one-hot encodings with conditional logic and Scikit-learn to support both statistical testing and machine learning.
- Performed statistical analysis using SciPy visualizations with Matplotlib and Seaborn, revealing that Q4 releases and Mature-rated games have significantly higher user ratings.
- Transformed a low-performing regression problem (~28%) into a classification task predicting 'loved' games (>4 rating) and trained a Random Forest Classifier model achieving 69% validation accuracy.

DATA SOURCES AND TOOLS

RAWG API: https://rawg.io/apidocs

Pandas Reference: https://pandas.pydata.org/docs/index.html

Scikit- Learn Reference: https://scikit-learn.org/

Seaborn Reference: https://seaborn.pydata.org/index.html

SciPy Reference: https://scipy.org/