LEUPHANA
UNIVERSITÄT LÜNEBURG

Video Game Ratings: Industry and Consumer Trend Analysis

Group 5
Pei I Wu
Abu Bakar Siddique
Eleanor Ruth Thomas
Farmand Bazdiditehrani
Jaykumar Himatbhai Savani
Mohammad Tabish Mohd Iqbal Momin

Table of Contents

1. Introduction

Pixels dance, ratings shift,
Trends weave through digital realms,
Clues to shape the game.

You can find the link to our notebook on GitHub.

In 2024, the global revenue from the video game industry was valued at 455 billion US Dollars (Clement, 2024). This is a fast-adapting industry, with new innovations being constantly developed (Goh, Al-Tabbaa, and Kahn, 2023). Therefore, insights into industry trends, user preferences, and game performance provide crucial advantages for developers and investors.

Our data appears to come from a video game tracking website. There exist several platforms for users to review their favorite games such as Steam, Metacritic, and so on. The array of data related to games and user responses provides the opportunity to analyze customer-game interactions. Based on this data, our analysis aims to deliver a Trend Analysis. We address the following research questions:

Trend Analysis

Genres	What are the most played genres? What are the most highly rated genres?
Teams	What companies dominate the market? Which gets the best reviews? On what specific genre do they focus?
Temporal	Based on release dates, how has the industry changed? How have the average ratings improved over the years? What genres piqued the most attention?
Rating	Are we able to predict which games will get high ratings? What are the key factors that influence this?

2. Dataset

Our data appears to be a snapshot dataset from a video game tracking website such as Grouvee or Backloggd. Each entry contains the data for a video game, with information such as release date, rating, number of plays, and reviews. Before cleaning, the dataset has 1512 entries and 14 variables.

Table 1: Overview of Dataset

	Unnamed: 0	Title	Release Date	Team	Rating	Times Listed	Number of Reviews	Genres	Summary	Reviews	Plays	Playing	Backlogs	Wishlist
0	0	Elden Ring	Feb 25, 2022	['Bandai Namco Entertainment', 'FromSoftware']	4.5	3.9K	3.9K	['Adventure', 'RPG']	Elden Ring is a fantasy, action and open world...	["The first playthrough of elden ring is one o...	17K	3.8K	4.6K	4.8K

In the following, to get a better feeling of the given dataset, each feature (column) is explained:

- **Unnamed 0:** Includes the index of each row, but it is redundant and does not add any value.
- **Title**: Name of the game
- **Release date**: The exact date that the game was released
- **Team**: The team that has developed the game

- **Rating**: Average user rating of the game out of 5
- **Times listed**: The number of listings (reviews)
- **Number of Reviews**: Similar to "Times Listed"
- **Genres**: Different genres associated with the games
- **Summary**: Brief description of the game
- **Reviews**: User-generated reviews in text format
- **Plays**: Total number of users who played the game
- **Playing**: Number of users currently playing the game
- **Backlogs**: Number of users who have bought the game but haven't played it (Users who plan to play the game later)
- **Wishlist**: Number of users who have added the game to their wishlist but haven't bought it.

As it is shown in Table 1, the first row of the dataset is presented. As mentioned above, the first column “Unnamed: 0” is redundant and is deleted in the next step (data cleaning). By looking at the other columns, we notice that the game Elden Ring was released officially on February 25, 2022, and was developed in collaboration with Bandai Namco Entertainment and FromSoftware teams. The average rating of this game is 4.5, which is no surprise since this game won the award of “Game of the Year” in 2022! The number of listings and reviews also shows that nearly 4 thousand users have shared their experience on this game! In the next column, the Genre of the game is noted in an array consisting of two genres of Adventure and RPG (Role-playing game). Summary and Reviews columns include a few sentences about the game, and finally the last four columns refer to the number of total players and buyers of this game.

3. Data Cleaning and Preparation

Data Types

Initially, all columns except “Rating” were stored as objects, and “Rating” was the only column with float64 (numerical) values. In some columns, numeric data was stored as objects (strings). For example, in the "Times Listed" column, we have values such as 3.9K which is equal to 3900. In order to be able to compute the average and do other descriptive statistics, we converted these to numeric columns. After this conversion, we ended up with 6 numerical columns.

Removing redundant columns

The first column only refers to indexes and it is of no value. Also, “Number of Reviews” contains the same values as the “Times Listed” column, and will not come in handy. Both were dropped.

Missing Values

The data has missing values in the “Rating”, “Team”, and “Summary” columns. These missing values are associated with games that are most recently released and therefore lack full entries. We removed the corresponding rows from our statistical analysis. The reason for dropping them instead of filling their values using the mean or median, is that 13 rows out of 1500 (less than 1% of the total) would not affect much.

Duplicated Data

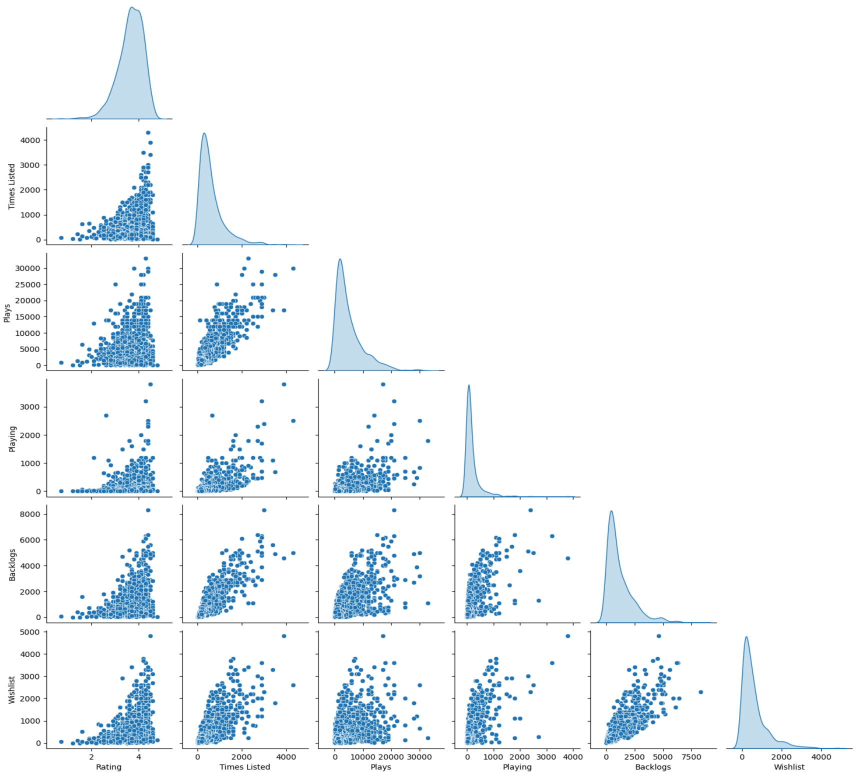
After applying the nunique method, we realized that there exist several duplicated data in our dataset. Deleting the duplicated rows only based on the repetitive titles could have misled us since some of the games have been developed in multiple versions over the years, and in the core they are different games. The dataset contained 382 duplicated rows. Therefore, we dropped these, leaving us with 1116 unique data entries to implement statistical analyses on them.

4. Statistical Analysis and Results

Exploratory Data Analysis (EDA)

In the next step, to understand the dataset better, we tried different approaches. We applied various statistical tests such as normality tests, chi-square test, and also examined the correlation between the features. Then, we made use of the principal component analysis and also examined the specific role of genres and teams (companies) in this dataset. Temporal analysis is also applied to understand trends over time better, and finally, we touched upon the algorithm of K-Means and cluster analysis.

Distributions, Normality Tests, ANOVA, and Chi-Square test

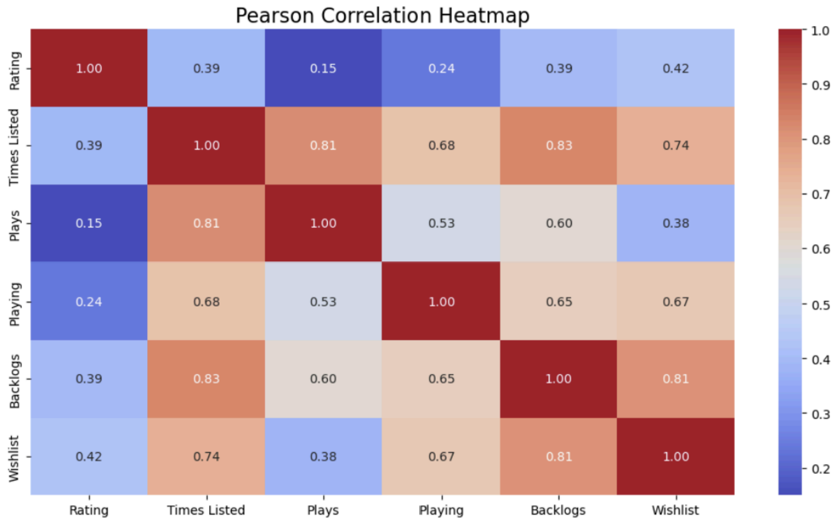


At first, using pair plot and box plot, we got to know whether the data points are scattered *normally* or not. After applying normality tests like *Shapiro-Wilk*, from the diagonal histograms above, you can see that most of the variables (particularly Times Listed, Plays, Playing, Backlogs, and Wishlist) are heavily right‑skewed rather than symmetric, so they are not normally distributed. The Rating variable might look less skewed than the others, but overall the data do not follow a bell‑shaped (Gaussian) distribution. Because of this, we employed the *log transformation* to deal with this problem.

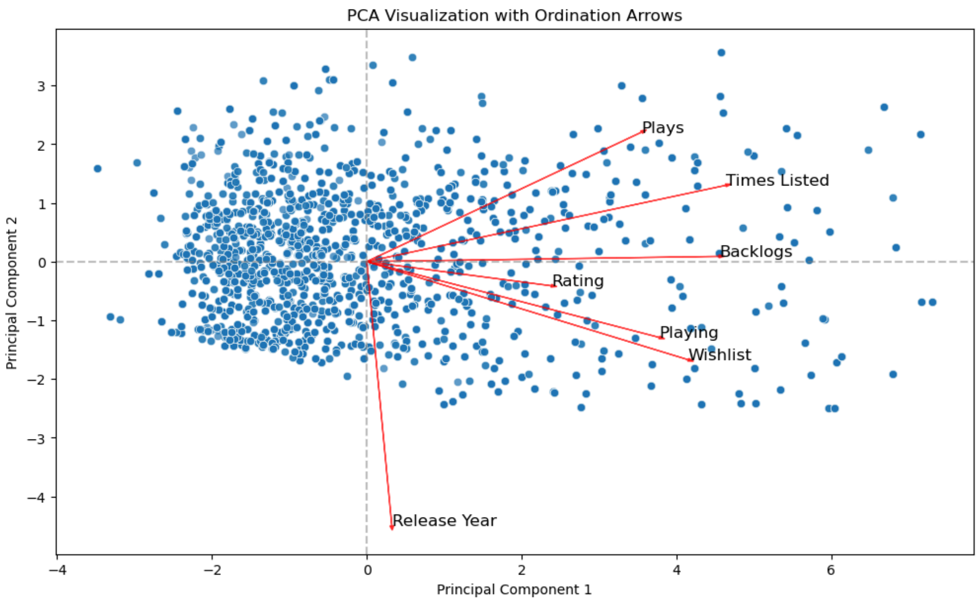
Applying Analysis of Variance *(ANOVA)* between genres and their ratings shows that since the p-value is extremely small (< 0.05), this indicates that there is a statistically significant difference in average ratings across different game genres. This means that at least one genre has a significantly different average rating compared to the others.
Also, the *chi-square test* between developer teams and the genres they have worked on shows that since the p-value is extremely small (< 0.05), the null hypothesis is rejected. This means that there is a statistically significant relationship between game teams and the genres they develop. In other words, certain teams are more likely to develop specific genres.

Correlations

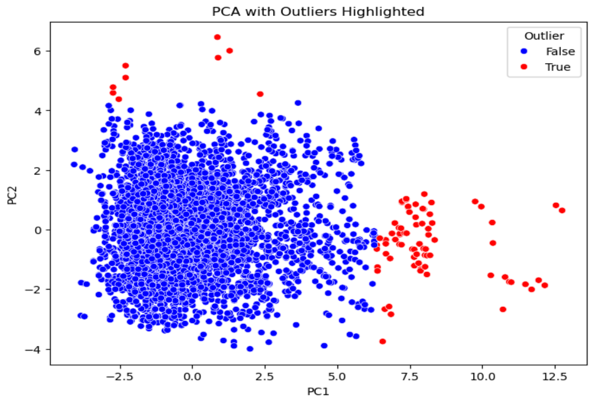
Using Spearman and Pearson's correlation coefficient and principle component analysis, we found that many of the count data are correlated, e.g. 'Plays', 'Times Listed', 'Backlog', 'Playing', and 'Wishlist' are all positively correlated. This suggests there is some redundancy within this data, which could all be considered as 'user engagement'.
Backlogs and Wishlist are both positively correlated with Rating, suggesting that if a game is highly rated, more people will want to play it. Also, the number of plays alone does not strongly determine a game's rating (weak correlation).



Principal Component Analysis



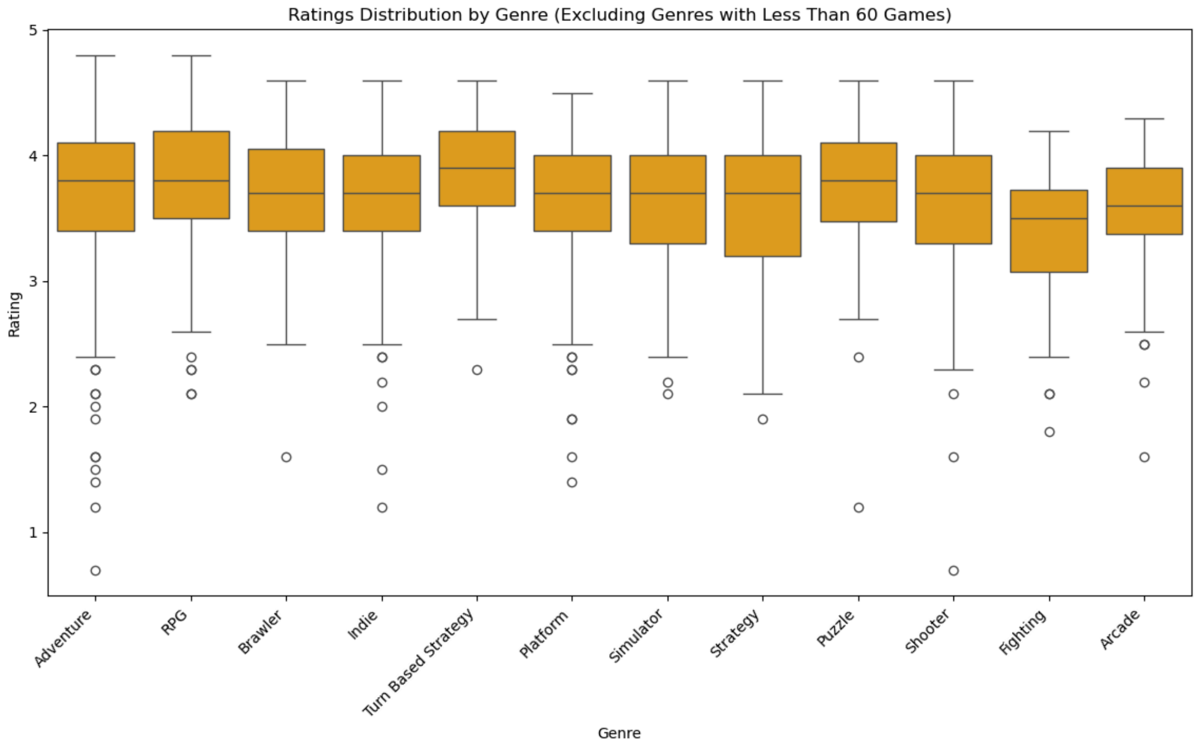
In the initial PCA, we were able to visually identify outliers. Identifying these outliers and comparing them to the overall data, the mean number of all user engagement and ratings were higher. This suggests there is a subcategory of games that have a very large user engagement and are likely ‘big-hit’ mainstream games.



Game Genre Analysis

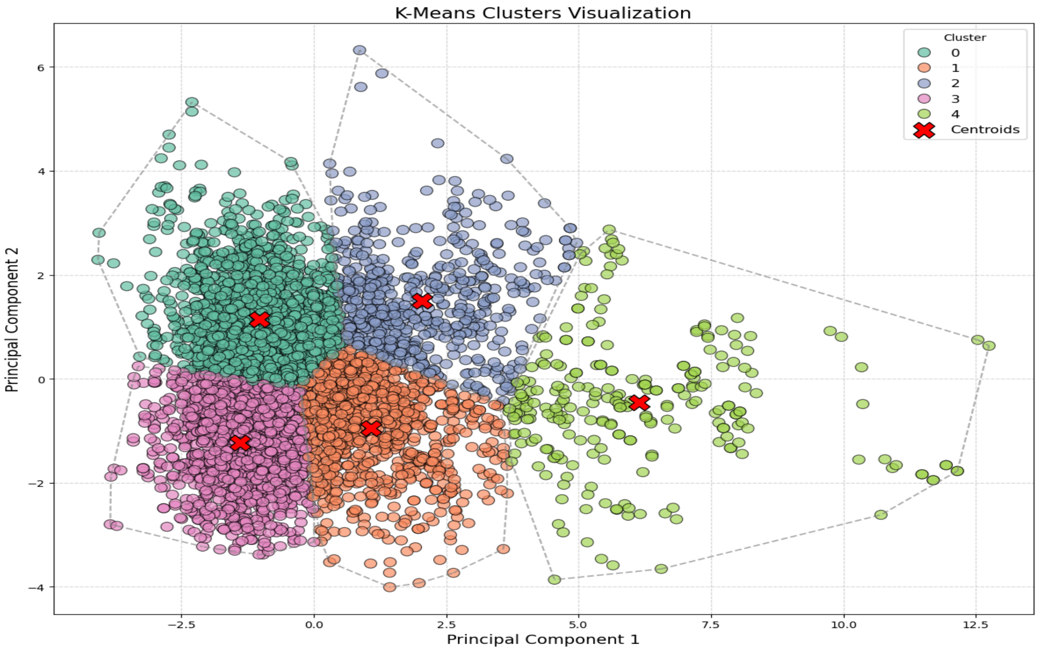
The boxplot illustrates the distribution of game ratings across various genres, excluding those with fewer than 60 games to ensure statistical reliability. Most genres have a median rating between 3.5 and 4.0, indicating that games generally receive favorable reviews across different categories. However, the spread of ratings varies, with some genres showing a wider range of scores, suggesting differences in player reception. A visual inspection suggests there is no significant difference between genre ratings.

Genres such as RPG and Puzzle tend to have more compact distributions, indicating that their ratings are relatively consistent, with fewer extreme values. On the other hand, genres like Fighting and Strategy show a wider IQR and long whiskers, indicating greater variability in ratings. This suggests that games in these genres receive more diverse feedback, with some titles being rated significantly higher or lower than others. Additionally, the presence of several outliers in most genres, particularly at the lower end of scale, indicates that while the majority of games receive moderate to high ratings, some titles have performed poorly.



Additionally, a cluster analysis was conducted to see if these genres are present in the numerical data. Based on the **elbow method**, **K=5** clusters were selected, as it provides a balance between well-defined clusters and avoids excessive fragmentation (see Appendix).

The **silhouette score for K=5 is 0.3640** (see Appendix), which indicates a moderate level of cluster separation. A silhouette score ranges from -1 to 1, where values closer to 1 suggest well-separated clusters, while values near 0 indicate overlapping clusters. The moderate score suggests that while the clusters have some separation, there is still a degree of overlap between them.



The results show clear differentiation between clusters based on their mean rating and principal component values. **Cluster 3** has the highest mean rating (4.219), indicating games in this cluster are generally well-received. Whereas **Clusters 1 and 2** have lower ratings, particularly Cluster 2 (3.48), suggesting these games may belong to less popular or lower-rated genres.

Cluster 3 has a significantly high PC1 value, representing games with distinct features contributing to high ratings. Based on the outliers found in the earlier PCA analysis, this may be games with very high user engagement. In contrast, **Cluster 1** has negative PC1 and PC2 values, suggesting a unique but possibly lower-rated game group.

Consider which genres are dominant in each cluster, we found three clusters where genre is dominant:
Cluster 3: RPG, Brawler, Shooter
Cluster 0: RPG, Indie, Platformer
Cluster 4: RPG (consistent high performance)

The cluster analysis of genre shows that genres like RPG, Shooter, and Brawler tend to receive higher ratings. In contrast, the games in Clusters 1 and 2 indicate groups with lower ratings.

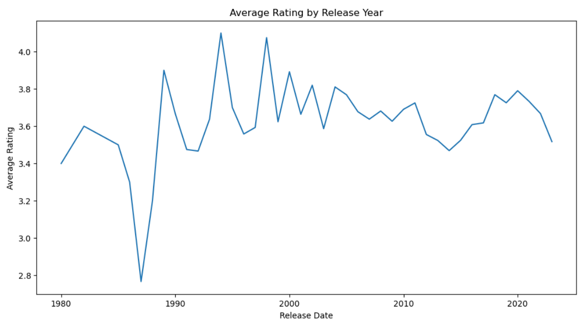
Developer Team Analysis

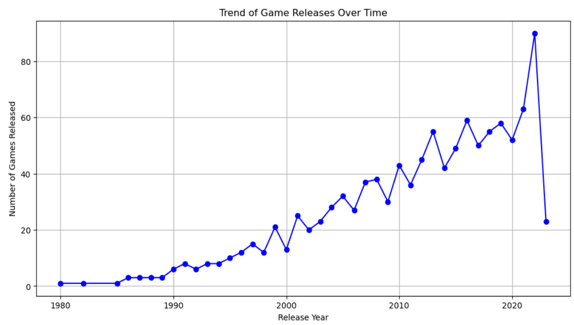
The Chi-Square test conducted to analyze the relationship between game development teams and genres yielded a statistically significant **p-value of less than 0.05,** indicating that game studios do not develop games across randomly but instead specialize in specific categories. Large AAA studios predominantly focus on high-performing genres such as FPS, RPG, and open-world Adventure games, leveraging their established expertise and audience base to maintain dominance in these areas through sequels and franchise-driven titles. On the other hand, independent developers demonstrate a greater degree of diversity, often targeting niche markets like Simulation, Strategy, Puzzle games, where innovation and experimental gameplay mechanics are more common. The test also revealed that certain studios have shifted towards different genres over time, likely in response to evolving market trends, player preferences and technological advancements. The presence of p-value (< 0.05) confirms that these patterns are not due to chance but rather strategic business and creative decisions.

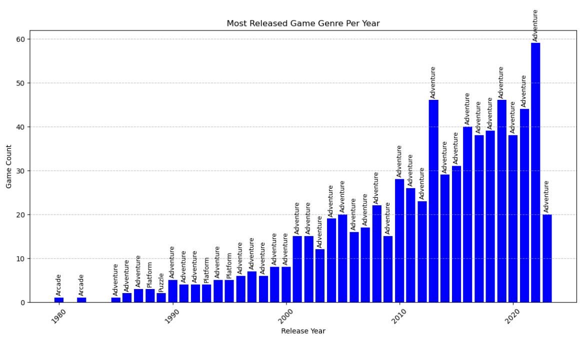
Temporal Analysis

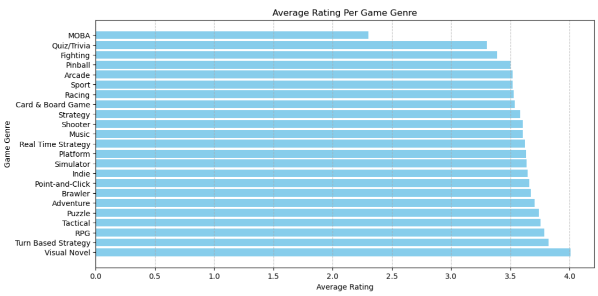
From the top-left plot, we can see that the average game rating was at its lowest around 1988. The top-right plot shows that the number of games released in 1988 was actually higher than in 1985, a year with a higher average rating. So, we know the low average score wasn't just because fewer games were released and they all happened to be unpopular.
Looking at the bottom-left plot, the most released game genre in 1988 was platformers. The bottom-right plot shows that platformers didn't have particularly low average ratings. So, it seems like the low scores in 1988 were more about the games themselves just not being well-received rather than the genre being a factor.

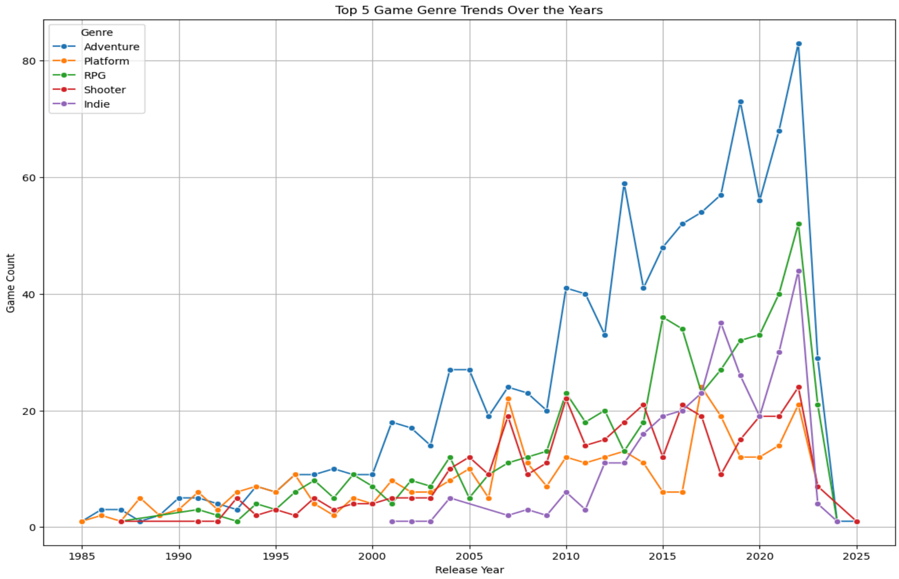
The final plot shows the release trends of the top five most common game genres (Adventure, Platform, RPG, Shooter, Indie) over the years. We can see that the overall trend of game releases is increasing, meaning that the gaming market has been gradually expanding, especially after 2000. Among these genres, adventure games have consistently had the highest number of releases. The plot also shows that Adventure, RPG, and Indie games have grown the fastest in the past decade, which may be related to market trends, technological advancements, and the rise of indie games.











Rating Predicting

Common ML algorithms such as Linear Regression, Logistic Regression, and Support Vector Machines were applied. However, their results were not insightful in comparison to the GLMs.

We used a GLM (Generalized Linear Model) to analyze how different variables affect game ratings. The variables we looked at include:

- The number of times a game was listed/reviewed (Times listed)
- Total number of users who have played the game (Plays)
- Number of users currently playing the game (Playing)
- Number of users who bought the game but haven't played it yet (Backlogs)
- Number of users who added the game to their wishlist but haven't bought it yet (Wishlist)
- Popularity

In GLM, the coefficient tells us the direction and strength of a variable's impact on ratings. If the coefficient is positive, the variable has a positive impact on the rating, meaning as the variable increases, the rating tends to go up. If the coefficient is negative, the variable has a negative impact, meaning as the variable increases, the rating tends to go down. The larger the absolute value of the coefficient, the stronger the influence of the variable.

The p-value ($P>|z|$) helps us determine whether a variable has a statistically significant effect on ratings. If $P \leq 0.05$, the effect is significant; if $P > 0.05$, the effect is not statistically significant.

Key Findings:

Times listed: The coef is **0.0006**, and **P < 0.001**, meaning the number of times a game appears on lists has a significant **positive impact** on its rating. This makes sense—highly rated games tend to get added to more lists.

Plays: The coef is **5.008e-06**, and **P = 0.777**, meaning the number of people who have played the game **does not have a significant impact** on its rating.

Playing: The coef is **-0.0001**, and **P = 0.002**, meaning the number of people currently playing the game has a significant **negative impact** on its rating. This suggests that as more people are actively playing a game, its rating might slightly drop.

Backlogs: The coef is **8.471e-05**, and **P = 0.001**, meaning the number of people who bought the game but haven't played it yet has a significant **positive impact** on the rating. The more people who own but haven't played the game, the higher the rating tends to be.
Wishlist: The coef is **9.124e-05**, and **P = 0.048**, meaning the number of people who added the game to their wishlist is **almost statistically significant**. This suggests a slight positive correlation between wishlist adds and ratings.
Popularity: The coef is **-4.936e-05**, and **P = 0.002**, meaning popularity has a significant **negative impact** on ratings. In other words, the more popular a game is, the more likely its rating might slightly decrease.
Overall, some variables, like Times listed and Backlogs, have a strong positive impact on ratings, while variables like Playing and Popularity have a slight negative impact. Others, like Plays and Wishlist, don't seem to have a clear effect.

Generalized Linear Model Regression Results

Dep. Variable:	Rating	**No. Observations:**	1116
Model:	GLM	**Df Residuals:**	1110
Model Family:	Gaussian	**Df Model:**	5
Link Function:	Identity	**Scale:**	0.22111
Method:	IRLS	**Log-Likelihood:**	-738.45
Date:	Thu, 27 Feb 2025	**Deviance:**	245.43
Time:	13:08:04	**Pearson chi2:**	245.
No. Iterations:	3	**Pseudo R-squ. (CS):**	0.2762
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	3.4396	0.023	152.850	0.000	3.395	3.484
Times_Listed	0.0006	7.35e-05	8.490	0.000	0.000	0.001
Plays	5.008e-06	1.76e-05	0.284	0.777	-2.96e-05	3.96e-05
Playing	-0.0001	4.56e-05	-3.047	0.002	-0.000	-4.96e-05
Backlogs	8.471e-05	2.44e-05	3.465	0.001	3.68e-05	0.000
Wishlist	9.124e-05	4.61e-05	1.979	0.048	8.95e-07	0.000
Popularity	-4.936e-05	1.61e-05	-3.072	0.002	-8.08e-05	-1.79e-05

5. Summary

Trend Analysis
Highly rated games tend to have more user engagement, but some lower-rated games still attract interest, perhaps due to franchise loyalty or unique dynamics. Newer releases that are backed by major studios often get high initial ratings, whereas older games maintain engagement.

Genres
Adventure games stand out as the most produced and played genre. However, there is no significant difference in rating between genres of game. Cluster analysis also reflected that the cluster with the lowest mean score was not linked to a specific genre. However, some genres tend to produce more consistent ratings, e.g. RPG and puzzle games. In contrast, others have a broader IQR, suggesting their reception is more variable, e.g. Fighting and Shooter.

Teams
Game developed teams do not randomly work on games but tend to focus on specific genres. There is a divide between large-scale developers who focus on high performing genres. Nintendo is by far the largest game developer team on the market.

Temporal
Although some years (likely those attached to very high performing games) show high ratings, generally there is no temporal trend in ratings. Indie games appear to have increased in number since the millennium and perform equally well to large developer teams. Therefore, although the market is dominated by larger companies, there remain investment opportunities in smaller companies.

Predicting Ratings
Although we were able to predict rating through generalized linear models, it is likely that the correlation in these models functions in the other direction: highly rated games likely appear on wishlists and backlogs *because* they are highly rated, rather than the reverse.

Limitations

Dataset Limitations
This data came from a specific subsection of the game consumers (those who use these websites). Therefore, it is possible that the data is not representative for the whole customer market. It also means that data on historical games (any released before the foundation of this website) display a survivorship bias towards games that have managed to remain played.

This data is also a database snapshot from a specific point. This gives us no historical data on ratings, meaning our analysis cannot reflect how ratings have changed over time. For example, we could not assess how the release of an update affects a game's performance.

We are also only given one average value for each game rating. This means we cannot look at any user-specific analysis, or look at the variance within ratings (e.g. is a game very divisive or do most ratings agree?). Additionally, a key factor in the video game industry is the fast-paced development of the field (Goh, Al-Tabbaa and Kahn, 2023). Variables such as new hardware and technological advances like cloud computing are likely to impact users' expectations of games. These variables are missing from our data.

As mentioned above, the features on this dataset could have been much better for analysis. In addition to the number of players, if we had access to the total hours of play, for instance, we could have derived insights based on that. There also exist other user engagement criteria that could help us achieve better and accurater results.

Methodological Limitations
We include an ANOVA comparing game genre and rating, but we realized it may be violating the test assumptions, as the data points were not independent, as one game can appear in multiple genres.

We attempted to analyze the content of the game reviews using a principle component analysis. We limited noise in the data by focusing on English adjectives used in the reviews. In a two-way PCA, the scales were not interpretable. When we increased the number of principle components to 3, the explained variance of each principle component was less than 0.005. Therefore, we decided there was too much variance in the results to be able to extract meaningful patterns from the analysis.

Although we included them, the silhouette scores in k-means clustering were only moderate, meaning the clusters could be clearer.

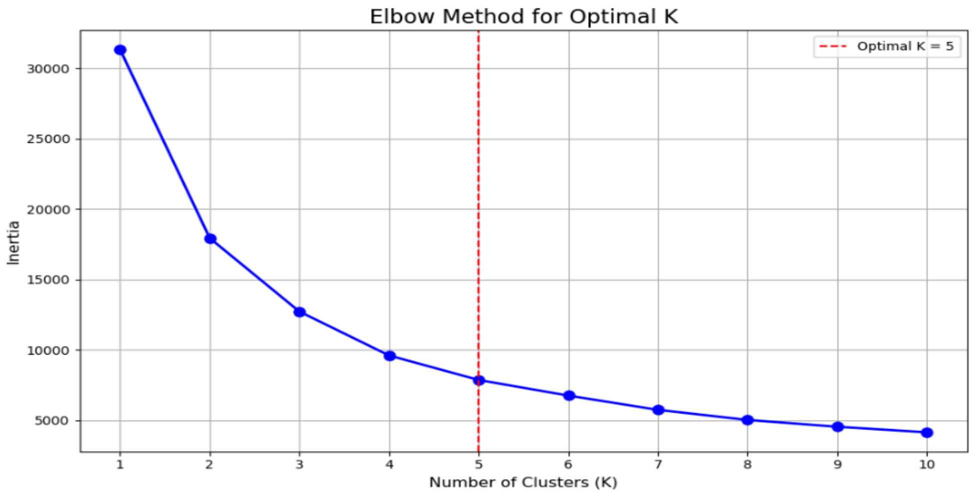
6. References

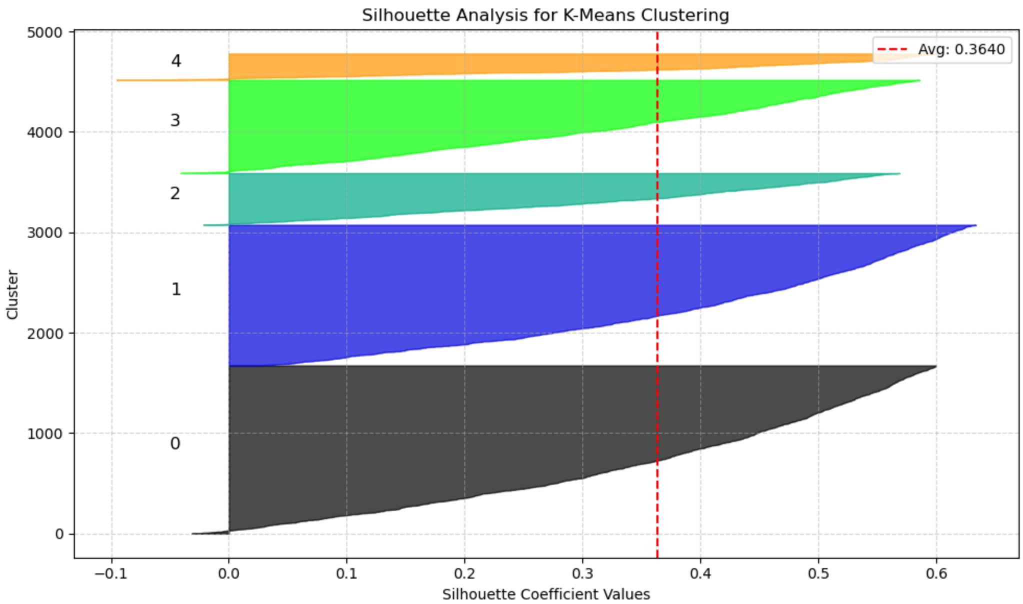
[1] Clement, J. (2024) Video game market revenue worldwide from 2019 to 2029. Available at: https://www.statista.com/forecasts/1344668/revenue-video-game-worldwide (Accessed: 27 February 2024).

[2] Goh, E., Al-Tabbaa, O. and Khan, Z. (2023) 'Unravelling the complexity of the Video Game Industry: An integrative framework and future research directions', Telematics and Information Reports, 12: 100100. Doi: https://doi.org/10.1016/j.teler.2023.100100

7. Appendix

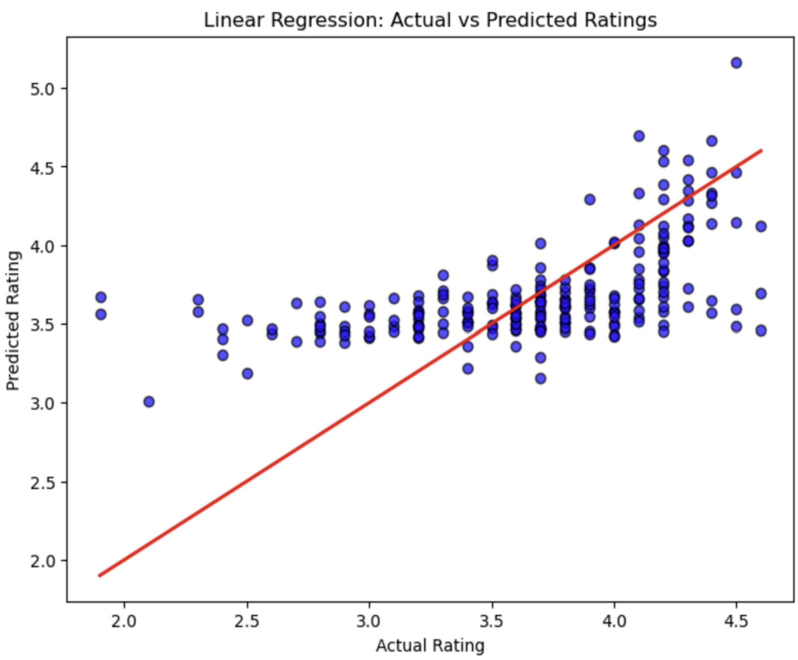
<u>Elbow Method and Silhouette Graphs from Cluster Analysis</u>





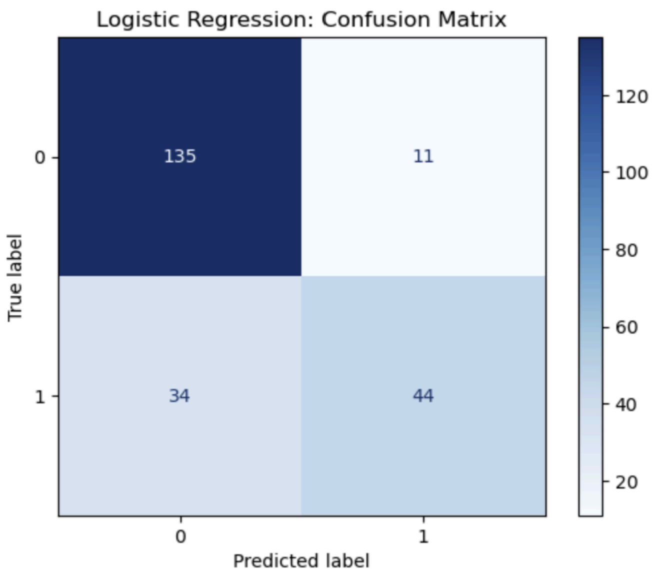
Common Machine Learning Algorithms Results:

- Linear Regression:

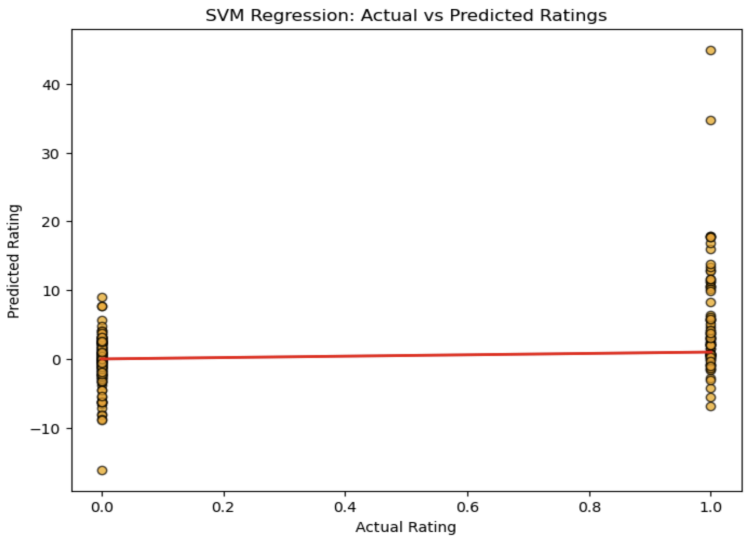


- Logistic Regression:

	precision	recall	f1-score	support
0	0.80	0.92	0.86	146
1	0.80	0.56	0.66	78
accuracy			0.80	224
macro avg	0.80	0.74	0.76	224
weighted avg	0.80	0.80	0.79	224



- Support Vector Machines:



<u>Generalized Linear Model result (AIC, Variance Inflation Factor - VIF):</u>

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                 Rating   No. Observations:                 1116
Model:                            GLM   Df Residuals:                     1110
Model Family:                Gaussian   Df Model:                            5
Link Function:               Identity   Scale:                         0.22111
Method:                          IRLS   Log-Likelihood:                -738.45
Date:                Thu, 27 Feb 2025   Deviance:                       245.43
Time:                        13:08:12   Pearson chi2:                     245.
No. Iterations:                     3   Pseudo R-squ. (CS):             0.2762
Covariance Type:            nonrobust
================================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept       3.4396      0.023    152.850      0.000       3.395       3.484
Times_Listed    0.0006   7.35e-05      8.490      0.000       0.000       0.001
Plays        5.008e-06   1.76e-05      0.284      0.777   -2.96e-05    3.96e-05
Playing        -0.0001   4.56e-05     -3.047      0.002      -0.000   -4.96e-05
Backlogs     8.471e-05   2.44e-05      3.465      0.001    3.68e-05       0.000
Wishlist     9.124e-05   4.61e-05      1.979      0.048    8.95e-07       0.000
Popularity  -4.936e-05   1.61e-05     -3.072      0.002   -8.08e-05   -1.79e-05
================================================================================
Selected Predictors: ['Times_Listed', 'Plays', 'Playing', 'Backlogs', 'Wishlist', 'Popularity']
Best AIC: 1488.901882854782
        Feature       VIF
0         const  2.555856
1  Times_Listed  9.012131
2         Plays       inf
3       Playing       inf
4      Backlogs       inf
5      Wishlist  4.661851
6    Popularity       inf
```