Predicting Video Games' Success

Chinh Ha, Inhoo Na, Yanaal Niyazi, Ziwei Yan

Introduction

The video game industry has evolved into a dynamic and expansive market since its inception in the 1970s, captivating millions of players of all ages worldwide with a diverse array of daily entertainment experiences. In recent years, as the market continues to flourish, the competition among game titles, genres, and platforms has become increasingly intense, which makes it more difficult for a video game to stand out or even sell enough copies to fund future developments. Therefore, it is important to predict the success of a game beforehand. This will enable game developers to use their money efficiently by tailoring their creations to meet market demands, allow publishers to optimize their portfolios, and empower investors to look for strategic opportunities that can maximize their profit.

In this paper, we will identify video game features that can help predict the success of a video game. The dataset we used is the Video Game Sales with Ratings data published on Kaggle by Rush Kirubi in 2016, which combined video game sales from the tracking website Vgchartz with corresponding ratings by staff and users from the game review website Metacritic. To measure the video game's success, we use the global sales and regional sales in North America, the European Union, Japan, and other parts of the world (i.e. Africa, Asia excluding Japan, Australia, Europe excluding the E.U. and South America) as the metrics. The factors we examine include the video game's genre, platform, publisher, ESRB rating (the age and content rating assigned by the Entertainment Software Rating Board to consumer video games in the United States and Canada), and user score (video game score given by Metacritic's subscribers).

The paper will start with a review of previous works on our central question of "can we predict the success of a video game based on its features such as genre, platform, publisher, and user ratings?" Then, it will explain the data cleaning process, statistical analysis, and applications of machine learning in predicting the success of video games along with data visualizations. Finally, we will draw conclusions based on the results.

Previous Works

This section presents existing papers and studies related to the topic of video games' success prediction.

1. *Critical Success Factors to Improve the Game Development Process from a Developer's Perspective*

Aleem et al. examined the success factors of video games from a developer's perspective by arranging organizational participants from different regions of the world to take part in a video game development process and take surveys. The study found that team configuration, management, testing, and programming practices can impact the success of a video game.

2. *A Survey of Marketing Management for the Video Games Industry in Turkey*

Through the survey conducted in the Turkish video game industry specifically, Scengun et al. drew the conclusion that promoting racial and sexual diversity in the game development team and creating a loyal customer base through marketing communication efforts are crucial to the success of a video game.

3. *Predicting Success Factors of Video Game Titles and Companies*

The study tried to find out the strategies, company organizations, and design decisions that can render video game titles successful by focusing on 144 games from 76 companies located in the European game industry. As a result, Pfau et al. found that genre, game engines, business models, and protagonist characteristics can highly influence a game's reception and economic accomplishment.

4. *Understanding mobile game success : a study of features related to acquisition, retention and monetization*

Moreira et al. considered the top 100 mobile games of the Google Play App Store and studied their number of downloads and revenues. The study showed that features like inviting friends and customizing the game can impact players' adoption of the game.

5. *Mobile games success and failure: mining the hidden factors*

In the study, Kerim and Genç investigated more than 17 thousand mobile games and discovered that in-app purchases, genre, number of supported languages, developer profile, and release month have a clear impact on the success of mobile games.

In conclusion, previous works regarding video games' success prediction span various facets, encompassing both the internal development processes and external market-oriented strategies. However, most of them only focused on a specific market or genre of video games instead of providing an overall understanding applicable across the entire spectrum of the video game industry. In our study, we will focus on the features of the game and consider video games from different genres, platforms, and developers in order to conclude factors that can be applied to predict success for a wider range of video games.

Methodology

After finding our dataset that includes data about video games and their ratings, genres, sales etc, we then proceeded to prepare and clean the data for a more interpretable and actionable dataset where we could produce insights from analysis on the dataset.

Data Preparation

1. *Invalid data within "User_Score"*

```
df['User_Score'].replace('tbd', np.nan, inplace=True)
df['User_Score'] = pd.to_numeric(df['User_Score'], errors='coerce')
```

We had to reform any cells within the dataset that contained data that was invalid such as 'tbd'(to be decided). As we realized there were cells that contained invalid data, we changed these cells into blank spots so that we could treat the user ratings uniformly.

This was done by writing a code that calculated the amount of missing values in each column of the dataframe and then prints out the count of missing values for the columns in a user-friendly format, by doing this we were able to clearly see which columns needed reforming as explained above. Additionally, we were able to write a code that replaced any cell within the "User_Score" column that had the input 'tbd' with 'np.nan' by using 'df.replace' which created the blank spots that we needed. We also converted the "User_Score" column to numeric values by using the Pandas library, more specifically 'pd.to_numeric'. This allowed for the analysis and preparation of the dataset to head in the right direction as we made sure the data was analyzable and in the right formats.

2. *Filling in the gaps with averages to create fairness*

```
for col in ['Critic_Score', 'User_Score', 'Critic_Count', 'User_Count']:
    df[col].fillna(df[col].median(), inplace=True)
```

While analyzing the dataset we came across the fact that there were also blank pieces of data in columns such as "Critic Score", "User Score", "Critic Count" and "User Count".

We used a loop that iterates over these columns and with the use of 'df[col].fillna' we were able to fill these blank spaces with the average of the other scores within the same column to basically 'predict' what that particular value/score would look like, instead of it being an empty piece of data. This is a common technique to handle missing data by inputting a central tendency measure. It holds the same concept as estimating a missing puzzle piece based on the surrounding pieces. By doing this we then have a more completed and fair dataset which can be used for further analysis with statistical and ML techniques.

3. *Cleaning up the categories in the dataset*

```
# Dropping rows with missing values in specified columns
df.dropna(subset=['Genre', 'Platform', 'Publisher', 'Developer', 'Rating'], inplace=True)
```

There were certain rows of data within important columns that have to be present for proper analysis such as "Genre" or the "Platform" the game is played on and we decided that if these pieces of data were incomplete/or partially missing it was best to completely remove these rows from our dataset to keep our analysis accurate and unflawed.

We did this by using 'df.dropna' on these certain columns so that any rows with missing values would be dropped from the dataframe and no longer included in our data awaiting analysis.

4. *Making categories suitable for analysis (machine learning algorithms)*

```
categorical_cols = ['Platform', 'Genre', 'Publisher', 'Developer', 'Rating']
df_encoded = pd.get_dummies(df, columns=categorical_cols)
```

One-hot encoding was a process we also included in our cleaning of the data so that certain columns that included categorical data would be converted to binary vectors (0's and 1's) so the data would be suitable for machine learning algorithms that require numerical input.

5. *Standardizing Sales figures*

```
# Normalizing sales figures
scaler = StandardScaler()
sales_columns = ['NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales']
df_encoded[sales_columns] = scaler.fit_transform(df_encoded[sales_columns])
```

We then decided to scale the sales figures to a common scale, for example how many copies were sold in North America or Europe. This was due to the fact that if all the sales were scaled to something such as the US Dollar or Euros, it would simply make the comparisons between sales of regions more interpretable.

This is done with the use of 'StandardScaler()' within our code, which allows for the scaling of all the sales through all columns.

Statistical Methods

1. *T-test for Action and Adventure Genre Critic Scores*

```
# T-Test
# Comparing Critic_Scores for Action and Adventure genres
action_scores = df[df['Genre'] == 'Action']['Critic_Score']
adventure_scores = df[df['Genre'] == 'Adventure']['Critic_Score']
t_stat, p_value = ttest_ind(action_scores, adventure_scores, nan_policy='omit')
print("T-Test for Action and Adventure Genre Critic Scores:")
print("T-Statistic:", t_stat, "P-Value:", p_value)
```

Inputs:
We inputted the critic scores from the Action and Adventure columns in our dataset by merging each genre and their critic scores into a variable name and then comparing both variables, aka 'action_scores' and 'adventure_scores'.

Test analysis:
With the use of the T-test we were hoping to find a correlation between the critic scores of these two genres, if the T-statistic were to be around the value of 2 or above, then we could confidently conclude that there is a significant difference between them and would be able to say that looking at the difference in critic scores is a good factor when comparing video games as a whole.

2. *Chi-square test for Platform and Genre*

```
# Chi-Square Test
# Relationship between Platform and Genre
platform_genre_table = pd.crosstab(df['Platform'], df['Genre'])
chi2, p, dof, expected = chi2_contingency(platform_genre_table)
print("Chi-Square Test for Platform and Genre:")
print("Chi2 Statistic:", chi2, "P-Value:", p)
```

Inputs:
For this test we inputted the data from platform and genre from our dataset by using the crosstab function from the pandas library.

Test analysis:
The chi-square test is something that we wanted to use to see how much the observed frequencies deviate from the expected frequencies in which we wanted both factors which were the platform and genre to be associated with each other and significantly correlated to the success of a game.

Machine Learning Methods

1. *Random Forest Classifier for Predicting Video Game Success for Genre, Platform, Publisher, Rating, and User Score*

```
df_encoded = pd.get_dummies(df_clean, columns=['Genre', 'Platform',
'Publisher', 'Rating'])
df_encoded['User_Score'] = pd.to_numeric(df_encoded['User_Score'],
errors='coerce')
feature_columns = [col for col in df_encoded.columns if col not in ['Name',
'Developer', 'Year_of_Release', 'NA_Sales', 'EU_Sales', 'JP_Sales',
'Other_Sales', 'Global_Sales', 'Success']]
features = df_encoded[feature_columns]
target = df_encoded['Success']
```

Inputs:
The Random Forest Classifier forecasts the 'Success' of video games, defined as worldwide sales above the median. "Genre,' 'Platform,' 'Publisher,' 'Rating,' and 'User_Score' are some of the features employed. These data points were chosen because we wanted to test them to see if all of them at once would impact a game's success.

Test Analysis:

```
rf = RandomForestClassifier(n_estimators=100, random_state=42)
rf.fit(X_train, y_train)
rf_predictions = rf.predict(X_test)
rf_accuracy = sum(rf_predictions == y_test) / len(y_test)
print("Random Forest Classifier Accuracy:", rf_accuracy)
```

Because of its capacity to handle both category and numerical data and its resistance to overfitting, the Random Forest Classifier is suitable for this task. To predict game success, the model is trained using these features. The number of trees (n_estimators) was changed to increase performance. The model is started with n_estimators=100 in the code, and its performance is assessed by its accuracy on the test set. To complete the study, add experiments in which n_estimators are adjusted to see how it impacts model correctness.

2. *KNeighborsClassifier for Genre, Platform, Publisher, Rating, and User Score*
   Inputs:
   We anticipated the 'Success' of video games using the same features as the Random Forest Classifier: 'Genre,' 'Platform,' 'Publisher,' 'Rating,' and 'User_Score'. These elements have been chosen based on their potential influence on game sales and popularity.

   Test Analysis:

```
knn_predictions = knn.predict(X_test)
knn_accuracy = sum(knn_predictions == y_test) / len(y_test)
print("K-Nearest Neighbors Classifier Accuracy:", knn_accuracy)
```

We use the KNeighborsClassifier since it perform well in classification tasks, in cases with non-linear connections between variables. The main variable is the number of neighbors (n_neighbors). The code's default option is n_neighbors=5. Varying this parameter is critical for understanding how it affects model performance since it controls how the classifier constructs the decision boundary based on the nearest points. To display a comprehensive experimental record, your code should contain trials with varying n_neighbors values, demonstrating how these differences impact the model's accuracy in predicting video game success.

3. *RandomForestClassifier for only game rating as in ESRB Game Rating (Age Group)*
   Inputs:

```
categorical_cols = ['Platform', 'Genre', 'Publisher', 'Developer', 'Rating']
df_encoded = pd.get_dummies(df, columns=categorical_cols)
X = df_encoded.filter(regex='Rating_.*')
y = df_encoded['Global_Sales']
```

In this code, we focus on predicting video game 'Success' based on the ESRB 'Rating'. We encode categorical variables like 'Platform,' 'Genre,' 'Publisher,' 'Developer,' and 'Rating' after cleaning and preparing the dataset. The Random Forest Classifier uses encoded 'Rating' columns as features, and the goal is the 'Global_Sales', which is binarized based on whether sales are above or below the median.
Test Analysis:

```
rf_classifier = RandomForestClassifier(random_state=42)
rf_classifier.fit(X_train, y_train)
```

```
y_pred = rf_classifier.predict(X_test)
```

In this scenario, the Random Forest Classifier is utilized to forecast game success based on the ESRB's 'Rating' function. This strategy is suited for dealing with worldwide sales that are binary (above or below the median). However, the code supplied does not show any modification in any RandomForestClassifier parameter to increase speed. Your code should include experiments in which parameters such as n_estimators, max_depth, or min_samples_split are modified, and the related changes in model performance metrics (accuracy, precision, recall, F1-score) are observed and documented.

Results

Statistics

1. *T-Test:*

   T-Statistic: (1.52); This value indicates the difference in means between the critic scores of both genres. A higher value would suggest a more significant difference, this usually is a value of 2 or above, which in this case is not satisfied. As a result, it's clear that the difference between the genres isn't significant in terms of having a high critic score.

   P-Value: (0.128); This means that there is a 12.8% chance that we would actually observe a difference between these two genres due to random variation. Since the p-value is not below 0.05 we do not have enough evidence to claim that the critic scores between these genres are significantly different.

   The result of the T-Test tells us that the critic scores between the genres Action and Adventure do not differ significantly. Therefore, when attempting to predict the success of a game, the difference between critic scores in these two specific genres is probably not a strong enough factor to make certain claims about their relationship.

2. *Chi-squared test:*

   Chi-square statistic: (3178.2040471801292); this value is a measure of how much the observed frequencies, which in this case is the combinations of platform and genre, the high Chi2 value suggests a substantial difference between actual and predicted frequencies if Platform and Genre were independent. With the P-Value of 0.0, virtually zero, indicates that this correlation is highly unlikely to have occurred by accident, demonstrating a statistically significant relationship between platform and genre selection in video games.

   P-Value: (0.0); this value of (0.0) suggests that the observed association between platform and genre is extremely unlikely to have occurred by chance and there is surely a correlation. The choice of platform and genre are not independent and there is a statistically significant relationship between them both as the P-Value is also less than 0.05.

The significant results found from this Chi-square test is the following; the platform a game is released on and its corresponding genre are definitely related as the P-Value indicates. This could then allow for such an analysis that certain genres may be more popular than others on specific platforms and consoles, which can highly boost and influence the success of that particular game, so the platform has to be chosen carefully when developing a game with a specific genre or vice versa. A great example is how games such as FIFA or NBA2K should be developed onto consoles such as PlayStation and Xbox, meanwhile games like Among Us should be targeted more towards mobile platforms in order to get the most successful results.

Machine learning

1. *Random Forest Classifier for Predicting Video Game Success for Genre, Platform, Publisher, Rating, and User Score*

   The accuracy of the Random Forest Classifier in predicting video game performance varied with the number of estimators used, revealing a complex link between game attributes and success. The accuracy was 0.7473 with 10 trees and increased to 0.7993 with 200 trees. This result shows us that, if we use a lot of factors like genre, user ratings, and etc then we can predict if a video game will be successful in terms of sales. When we use a more detailed model, it gets better at predicting. This helps answer our main question, showing that understanding different parts of a game, like what type it is or who publishes it, can really help guess if it'll do well in the market.

2. *KNeighborsClassifier for Genre, Platform, Publisher, Rating, and User Score*

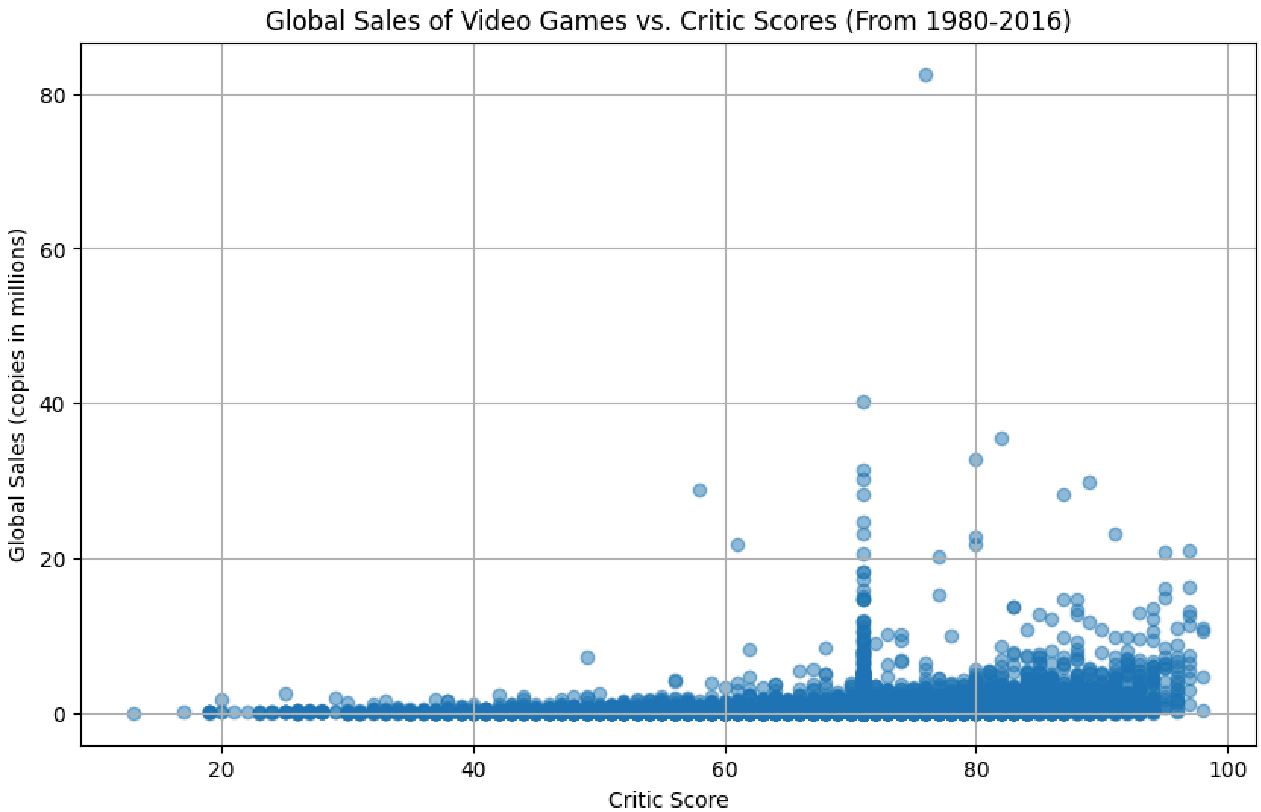
   The observed trend in the performance of the K-Nearest Neighbors Classifier, with accuracy peaking at 0.6777 with 15 neighbors, indicates a moderate but favorable association between the chosen characteristics and game success. This suggests that, while KNN may predict video game success to some extent, its performance falls short of that of more complicated models such as Random Forest. The rise in accuracy with additional neighbors (up to a point) highlights K&N's potential in exposing how many game elements such as genre, platform, publisher, rating, and user score may interact to impact success. However, even with the low accuracy, it proves a different perspective in predictive analysis, showing the viability of employing machine learning to predict video game success.

3. *RandomForestClassifier for only game rating as in ESRB Game Rating (Age Group)*

   When used to forecast video game success purely on ESRB Game Rating (Age Group), the Random Forest Classifier consistently produced an accuracy of 0.5183 across varying numbers of trees (n_estimators). This consistent accuracy, regardless of the increase from 10 to 200 trees, implies that the ESRB Game Rating alone may not be a good indicator of video game success. However, even if it's not a good indicator, it is critical to comprehend the multidimensional nature of game success prediction. It will tell us that, while ESRB ratings give some information, they are inadequate to reliably predict success on their own. This finding will contribute to our key topic by emphasizing the significance of a more diversified collection of attributes for better prediction.

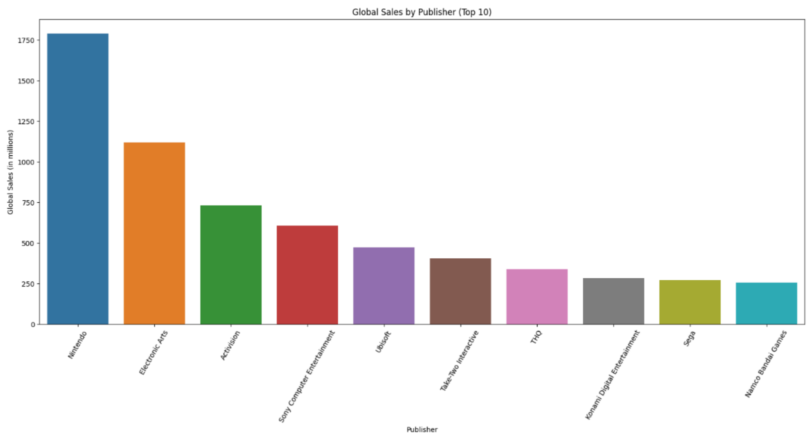
Visualizations

1. Critic Scores vs. Sales Scatter Plot



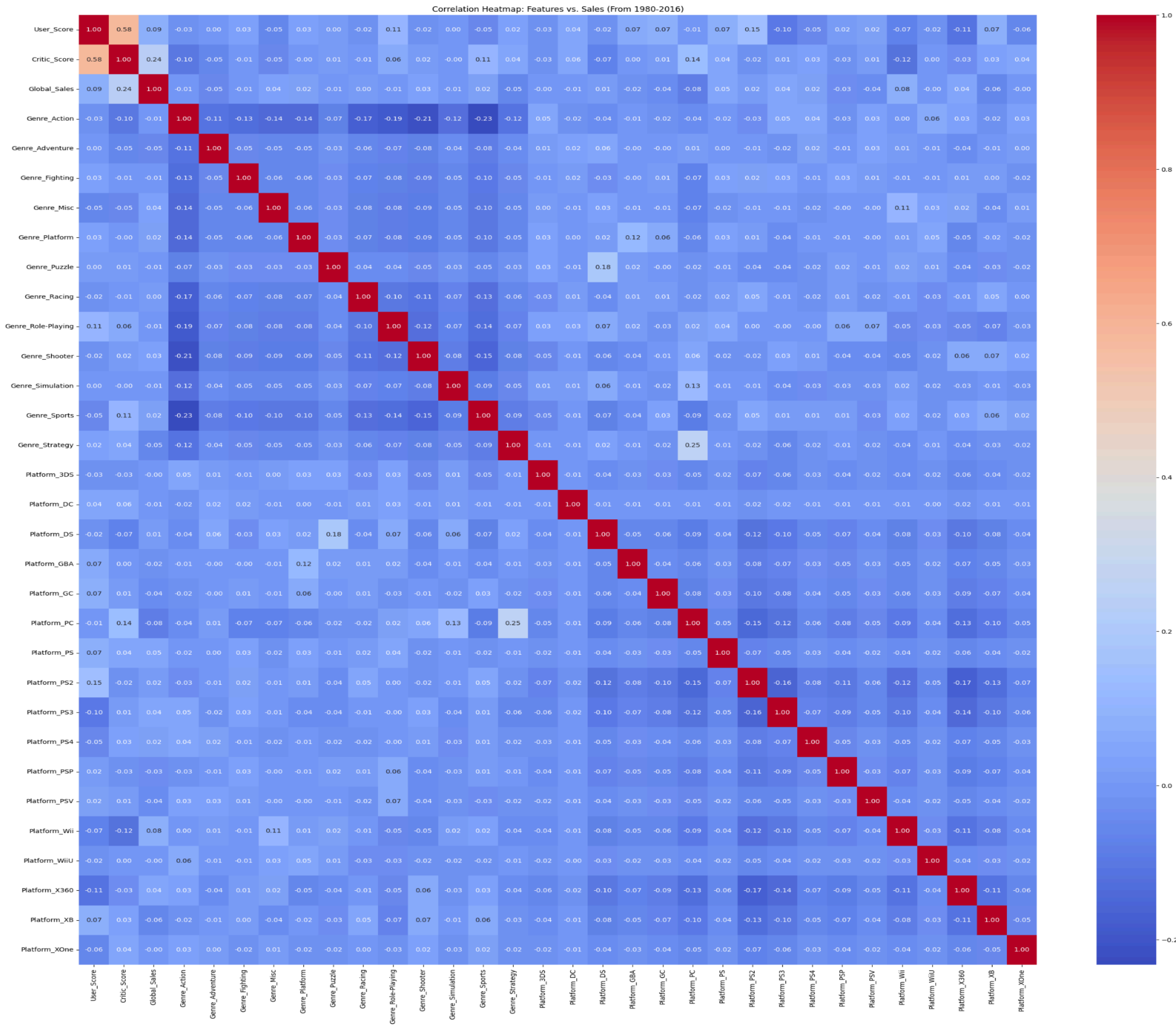
This is a scatter plot for a graph that takes in the critic score as the independent variable and using the global Sales that are counted per copies in millions. As we can see in the graph there seems to be a huge correlation where if the rating is higher from the critic then the sales would increase more.

2. User Ratings vs. Sales Bar Chart



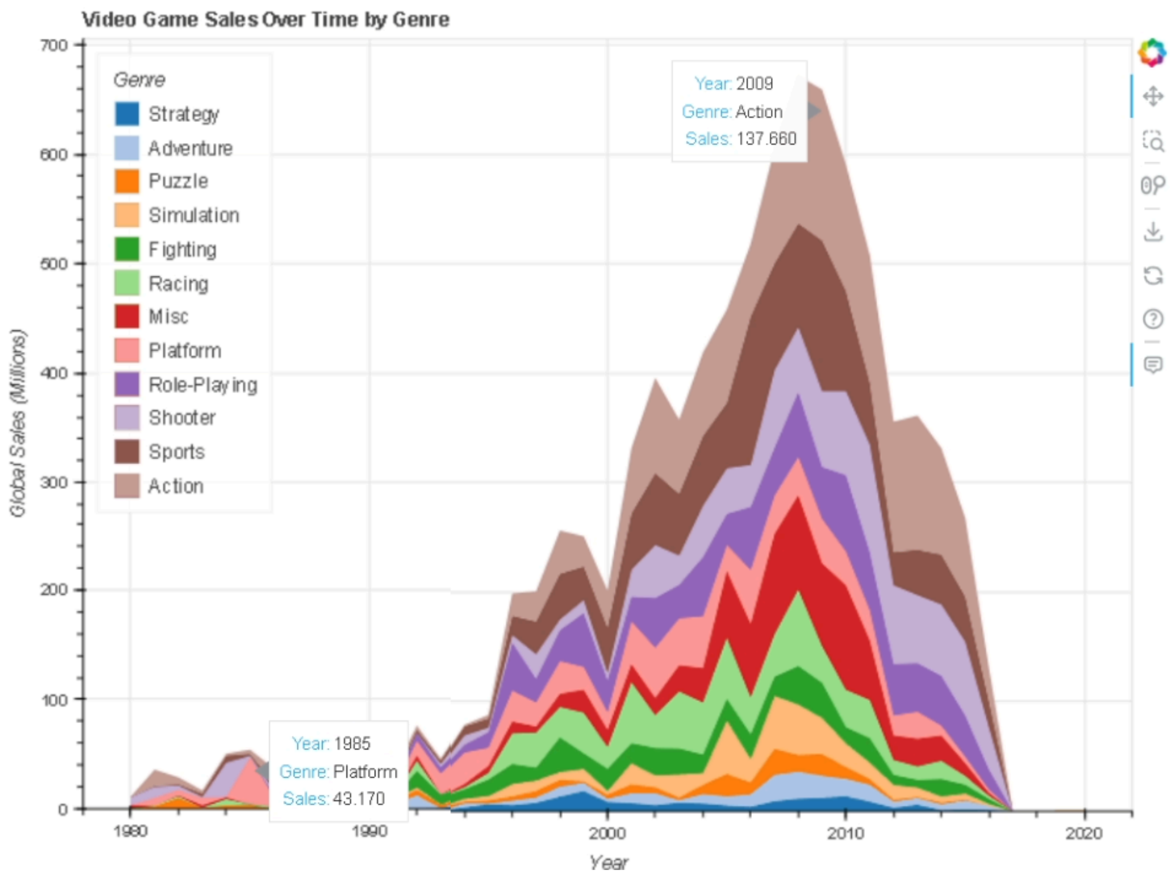
This is a bar graph that takes in the top 10 Publisher companies as the independent variable and uses the global sales from that company that are counted per copies in millions. As we can see in this graph there seems to be a huge correlation where depending on what company the game is released from the sales would increase more. For example if Nintendo releases a new game, while Ubisoft also releases a new game, then Nintendo would most likely have better global sales than Ubisoft.

3. Features vs. Sales Seaborn Heatmap



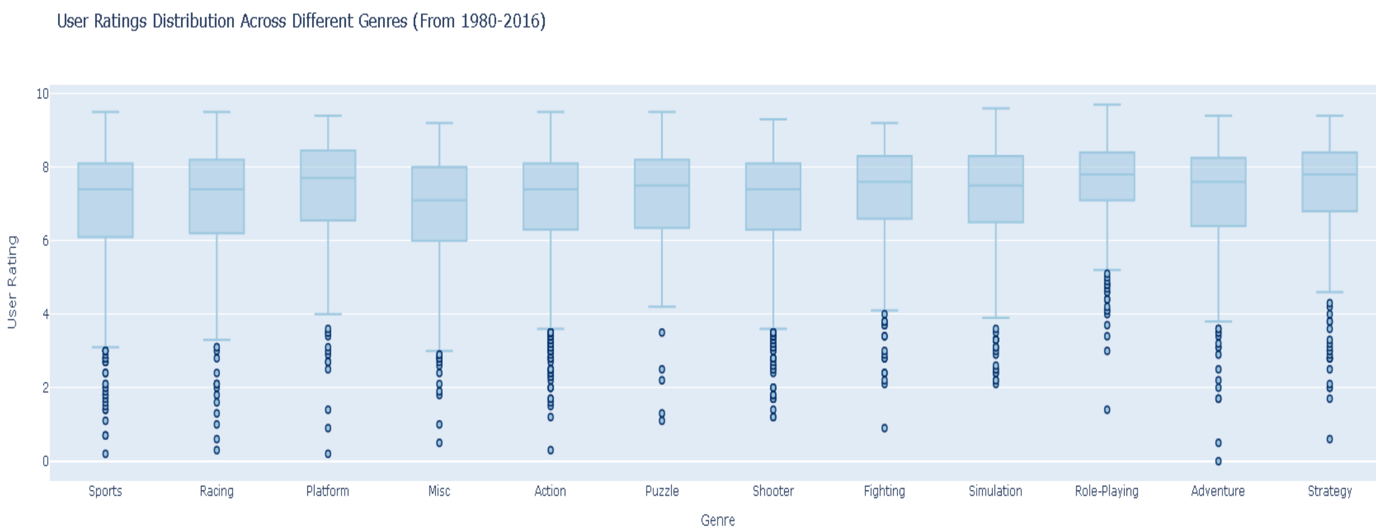
The correlation matrix's heatmap, which depicts the links between key elements of video games and their global sales. This image shows how several qualities link with sales in a color-coded manner, with warmer colors showing larger positive associations. However, as one can see, the heatmap has very few warm colors. The only thing that we can derive from this heatmap is that there is a bit of positive correlation between the strategy genre and pc platform, and global sales with critic scores. Everything else is too minimal or does not apply (like genre to genre correlation).

4. Stacked Area Chart for Video Game Globals Sales over time by Genre



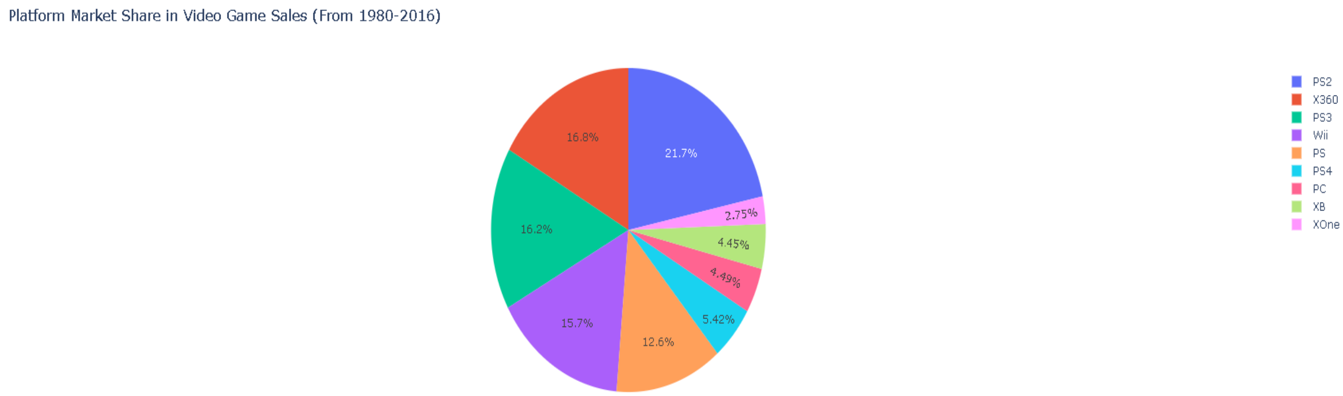
This is a stacked area chart that was generated by ChatGPT-3, that depicts the correlation between the genre of a video game and its global sales through 1980-2016 (ChatGPT-3). To read the stacked area chart, the height of the graph, the amount of sales there is in a game genre. The order of the genre does not matter too and they could be reordered and give us the same result. However for this graph, we decide to order each genre from least to most sales in total. As we can see in the chart it seem like platformer game did well back then in 1985, but it transition to action game being the top genre that is gaining the most sales out of all the games and the action genre seem to dominate the video games global sales in general while the sales are at its peak and at it's low too from 1993 to present day.

5. User Rating Across Different Genres Box Plot Plotly



The plot of the box Visualizing user ratings across various video game genres is an important analytical method for tackling the topic of forecasting video game performance based on factors such as genre. This graph depicts the distribution and variability of user ratings across genres, highlighting patterns and outliers. In this case though, because all of the statistics (mean, range, quartiles) are similar, this plot does not give us that much information. We cannot really derive much from the information given in this plot.

6. Global Sales for each Platform Pi Chart



This is a pi chart to clearly depict the correlation between the platform of a video game and its global sales through 1980-2016. As we can see the sales between Playstation 2, XBOX360, Playstation 3, and Wii are close in sales, however we can see that at the time PC sales in video games were low. Therefore if you were around in the time period of 1980 to 2016, then you would be more likely to sell better with consoles compared to PC.

Conclusions

In conclusion, our research into predicting the success of a video game based on its features such as genre, platform, publisher, and user ratings provided good results. The statistical methods, which were the t-tests, tell us that certain aspects, such as genre, have less of an impact in gaining high critic scores. However, in the Chi-squared test, the Chi2 statistic and P-Value reveal a significant association between the platform and genre of a video game. This means that some genres are more popular or perform better on various platforms. Understanding this link can assist in forecasting a game's success since it aids developers and marketers in selecting the best genre and platform combination. For example, if we are making a simple or quick strategy game like Clash of Clans, it may perform well on mobile, compared to putting the game on a console or PC. This Chi-square test finding may also collaborate with Moreira et al, "Understanding mobile game success: a study of features related to acquisition, retention and monetization" and Kerim and Genç "Mobile games success and failure: mining the hidden factors" since the articles focus on the game design of popular mobiles games that will attract consumer to play their game more, where our finding shows what genres will show the type of genre mobiles user would enjoy in their game.

The machine learning method also helps us see how a game in the past successes, notably the Random Forest Classifier, which revealed that a complicated model incorporating several characteristics may accurately predict game success. The model's accuracy varied with different parameters, emphasizing the complicated link between these qualities and game success. However, depending exclusively on the ESRB Game Rating was less effective, emphasizing the importance of a multifaceted approach to prediction. The data support the notion that game success is a multifaceted term driven by a variety of factors.

Last but not least, the visualization gives us an indicator of seeing if there is a positive correlation, where we can see variables just as the user rating, critic, rating usually equal to a large sales of video game copies, in the scatter plot and bar plot. We also see that certain genre and game companies usually have more sales compared to the other genre and game companies. The visualization helps us clearly see all of the dataset more clearly and easier too and helps us determine what data variables have a positive correlation with the global sales and also what didn't correlate as much with the global sales, for example with the bar chart that see the relationship between the game company with the globals sales they have made, Nintendo seem to have a huge amount of game sales compared to the other game company, however base on the heatmap, it tells us that the relationship between game company with video game sales is low, which tell us that Nintendo might be a outlier and the game company does not one hundred percent tell us if a game will do well.

While we have gotten very useful information and seeing what and what don't collaborate with a game's success in terms of sales, the investigation might be broadened by including additional variables such as how much they budgeted their game by ads, customer evaluations, or etc which have been emphasized in earlier research but were not included in our dataset. This comprehensive approach, which included both internal development elements and external market dynamics, has the potential to provide a more thorough understanding of what drives video game success. Our investigation of video game success taught us about the complexity of the elements that impact a game's market performance. We hope the idea of more datasets about the game will be created since different trends affect the game market and we will see variables, providing a broader, more nuanced understanding of the factors that contribute to video game success in the present day.

Reference

Aleem, S., Capretz, L.F., Ahmed, F.: Critical success factors to improve the game development process from a developer's perspective. J. Comput. Sci. Technol. 31(5), 925-950 (2016)

Kerim, A., Genç, B. Mobile games success and failure: mining the hidden factors. Neural Comput & Applic (2022). https://doi.org/10.1007/s00521-022-07154-z.

Moreira, A.V., Filho, V.V., & Ramalho, G. (2014). Understanding mobile game success : a study of features related to acquisition , retention and monetization.

OpenAI. "ChatGPT-3." 2023, www.openai.com. Accessed 12/5/2023.

Pfau, Johannes & Debus, Michael & Juul, Jesper & Hammar, Emil & Canossa, Alessandro & El-Nasr, Magy. (2022). Predicting Success Factors of Video Game Titles and Companies. 10.1007/978-3-031-20212-4_22.

Şengün, Sercan. (2018). Chapter 8.1: A Survey of Marketing Management for the Video games Industry in Turkey. 10.1108/978-1-78714-557-320181024.