****

FINAL REPORT

# Game Genre Analysis & Prediction

Team 12:

Ishan Subodh Joshi

Kunal Singh Pundir

**Introduction**

**Problem Statement:**

Understanding and predicting game genre trends is pivotal for developers, marketers, and investors in the rapidly evolving video game industry. The video game market is characterized by diverse genres, each appealing to different player demographics and preferences. However, the industry lacks a comprehensive analysis tool that can accurately analyze the current trends and predict future shifts in genre popularity. This gap leads to challenges in decision-making, investment, and development strategies for stakeholders.

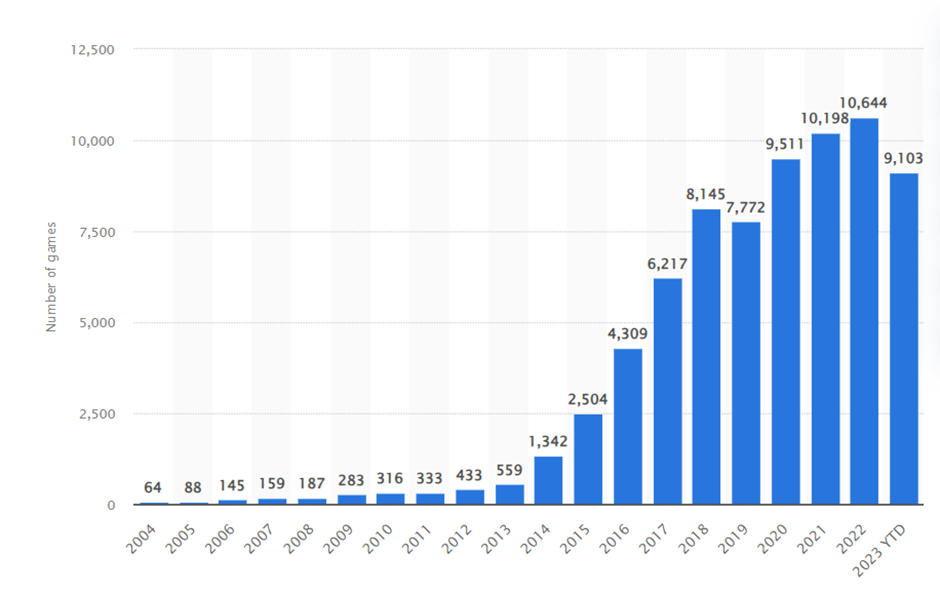
The primary issue is the absence of a data-driven approach to analyze the patterns and factors influencing the popularity and success of various video game genres. Traditional methods often rely on subjective assessments or historical sales data, which do not adequately capture the dynamic nature of player preferences and market trends. Additionally, the influence of social media, online reviews, and gaming platforms on genre popularity is poorly understood.

Moreover, the industry faces difficulty in predicting how emerging technologies (like VR and AR), new gaming platforms, and changing consumer behaviors will impact the popularity of different genres. This uncertainty complicates strategic planning for game development and marketing.

Our project aims to analyze historical data, the price of the game, the game description, game tags, and other relevant metrics to provide insights into current genre trends and predict future shifts in the video game market. This approach will enable stakeholders to make more informed decisions, better align their strategies with market dynamics, and potentially identify untapped opportunities in the video game industry.

**Background:**

The digital gaming landscape has undergone a revolutionary transformation with the advent of online platforms, and Steam stands at the forefront as a behemoth in the PC gaming realm. Boasting millions of users and an extensive library of games, Steam offers an unparalleled gaming experience.



**Source:** https://www.statista.com/statistics/552623/number-games-released-steam/

However, this abundance of choices challenges users to discover games that align with their unique preferences. Traditional recommendation systems often fall short in providing personalized suggestions, necessitating the application of advanced data science methodologies to harness the wealth of user interaction data available on the platform.

**Motivation:**

The digital age has revolutionized how we interact with entertainment, particularly in video gaming. As a leading digital distribution platform, Steam stands at the forefront of this revolution, boasting 120 million monthly active users and 62.6 million daily users. This staggering user base is a testament to the platform's success and reflects the diverse and ever-growing global gaming community.

The exponential growth of Steam's user base has resulted in an equally extensive and varied game catalog. While this diversity offers gamers a plethora of choices, it also presents a significant challenge: navigating through an overwhelming sea of options to find games that align with their preferences. The current recommendation systems, largely generic in their approach, fall short in addressing individual users' unique tastes and gaming patterns. This gap underscores the need for a more sophisticated, personalized recommendation system.

Our project is driven by the ambition to bridge this gap. By harnessing the vast amounts of data generated by user interactions on Steam—ranging from playtime statistics, purchase history, and user reviews to forum discussions—we aim to develop a recommendation system that is far more attuned to individual preferences. This system will not only consider the genres that users have historically preferred but will also intelligently identify emerging patterns in gaming behavior, potentially introducing users to new genres and experiences they might enjoy.

Furthermore, the project explores the dynamic relationship between gaming trends and external factors such as seasonal releases, gaming events, and social media influence. Understanding these correlations can significantly enhance our predictions' accuracy, offering users a reflection of their current preferences and a window into evolving trends that could shape their future gaming experience.

By achieving this, our project goes beyond the realm of simple entertainment; it steps into the domain of creating a more engaging, rewarding, and personalized gaming community on Steam. This approach has the potential to not only heighten the user experience but also offer valuable insights to game developers and marketers, enabling them to tailor their offerings more effectively to meet the desires of their audience.

Ultimately, our project aspires to redefine how gamers interact with Steam's vast catalog. It's about creating a system that understands the nuances of individual gaming identities, respects the diversity of player preferences, and fosters a more connected and satisfied gaming community.

**Goal:**

The primary goal of this project is to design and implement an intelligent and personalized suggestion model for Steam, leveraging the power of data science and machine learning. By analyzing user behaviors, game metadata, and other relevant features, the system aims to understand the nuanced preferences of individual gamers. The endgame is a recommendation algorithm that accurately predicts a user's gaming inclinations and introduces serendipitous discoveries. In achieving this goal, the project aspires to improve user satisfaction on the Steam platform and the broader discourse on the efficacy of data-driven approaches in enhancing recommendation systems for diverse online platforms.

**Methodology**

**1. Data Collection and Preparation**

• Sourced data from Steam, focusing on various attributes of video games.

• Loaded and read the dataset into a data processing environment for analysis.

**2. Data Cleaning and Preprocessing**

• Addressed null values in key columns, ensuring data quality and integrity.

• Converted date-related data to appropriate datetime formats for time-based analysis.

• Performed text data cleaning in several columns to make the data usable for analysis.

• Dropped columns with a high percentage of missing values to streamline the dataset.

• Normalized data formats, especially in text fields, for consistency.

**3. Exploratory Data Analysis (EDA)**

• Conducted descriptive statistics to summarize the dataset’s key characteristics.

• Utilized data visualization to uncover patterns and trends in the data.

• Performed correlation analysis to explore relationships between different game attributes.

**4. In-Depth Analysis**

• Analyzed trends such as the most downloaded and most expensive games.

• Compared free versus paid games in terms of average downloads.

• Investigated user preferences like language support in relation to game popularity.

• Assessed the most used operating systems for games and their impact on popularity.

• Evaluated games' ratings and their correlation with download figures.

**Dataset Description**

Each column provides detailed and specific information about the games, contributing to a comprehensive understanding of each title in the dataset.

**Dataset Link:** [**https://www.kaggle.com/datasets/fronkongames/steam-games-dataset**](https://www.kaggle.com/datasets/fronkongames/steam-games-dataset)

The columns in our dataset each provide specific information about the games on Steam. Following is the detailed description of each dataset:

Each column provides detailed and specific information about the games, contributing to a comprehensive understanding of each title in the dataset.

The columns in our dataset each provide specific information about the games on Steam. Following is the detailed description of each dataset:

**1. AppID:** Unique identifier for each game on Steam.

**2. Name:** Title of the game.

**3. Release Date**: When the game was made available on Steam.

**4. Estimated Owners:** The approximate number of people who own the game.

**5. Peak CCU (Concurrent Users):** The highest number of users playing the game simultaneously.

**6. Required Age:** Minimum age recommended to play the game.

**7. Price:** Cost of the game.

**8. DLC Count:** Number of downloadable content items available for the game.

**9. About the Game:** Brief description or synopsis of the game.

**10. Supported Languages:** Languages available for the game interface and subtitles.

**11. Full Audio Languages:** Languages available in the game's audio.

**12. Reviews:** User reviews of the game.

**13. Header Image:** Promotional image or banner for the game.

**14. Website:** Official website of the game.

**15. Support URL:** Link to game support or help page.

**16. Support Email:** Contact email for game support.

**17. Windows/Mac/Linux:** Compatibility with these operating systems.

**18. Metacritic Score:** The game's score on Metacritic.

**19. Metacritic URL:** Link to the game's Metacritic page.

**20. User Score:** User-generated score for the game.

**21. Positive/Negative:** Counts of positive and negative reviews.

**22. Score Rank:** Ranking of the game based on scores.

**23. Achievements:** List of in-game achievements.

**24. Recommendations:** Endorsements or recommendations for the game.

**25. Notes:** Additional notes or comments about the game.

**26. Average Playtime (Forever/Two Weeks):** Average time spent playing the game.

**27. Median Playtime (Forever/Two Weeks):** Median playtime statistics.

**28. Developers/Publishers:** Companies that developed and published the game.

**29. Categories:** Classification of the game (e.g., single-player, multiplayer).

**30. Genres:** Genre of the game (e.g., action, strategy).

**31. Tags:** Keywords associated with the game.

**32. Screenshots:** Images capturing gameplay or game scenes.

**33. Movies:** Promotional or gameplay videos.

**Result and Analysis**

**Data Collection & Data Preprocessing:**

In data collection, information is gathered from various sources, ensuring it's representative and relevant to the research question. During data pre-processing, this raw data is cleaned and transformed; missing values are addressed, data types are standardized, and anomalies are corrected. The goal is to refine the dataset into a structured, consistent format suitable for analysis, often involving feature extraction and normalization to aid in meaningful data interpretation and model performance.

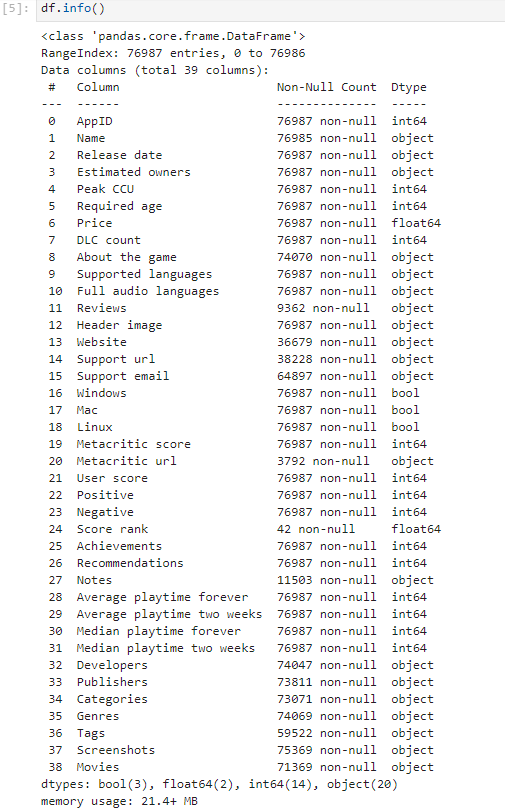
**Data Collection**

**Source of Data:**

The dataset was compiled from Steam, a comprehensive digital distribution platform for video games. The data encompasses various features related to video games listed on the platform.

**Scope of Data:**

The dataset includes 76,987 entries, each representing a unique video game. The data spans a wide range of information fields, from basic game details to user interaction metrics.



**Data Attributes:**

The dataset comprises 39 columns covering the following attributes:

* Basic game information such as AppID, Name, and Release date.
* Ownership and user interaction metrics like Estimated owners, Peak concurrent users (CCU), and playtime statistics.
* Game features include Required age, Price, and DLC count.
* Content descriptions are given in About the Game.
* Language support is detailed in Supported languages and Full audio languages.
* Media content like Screenshots and Movies URLs.
* Game categorizations are in categories, genres, and tags.
* Developer and publisher information.
* User ratings include positive and negative counts and Metacritic scores.

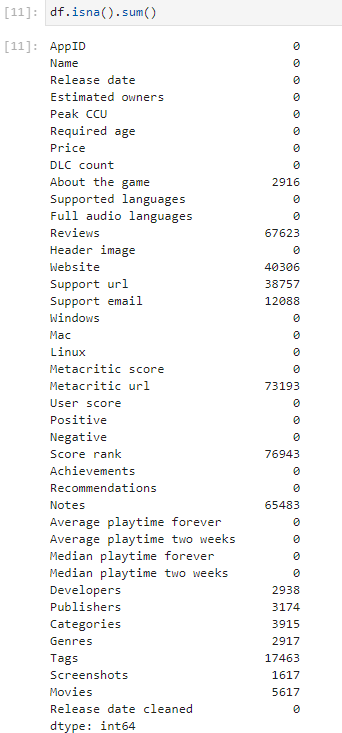
**Data Preprocessing**

**Data Cleaning Objectives:**

The preprocessing aimed to prepare the dataset for robust analysis by ensuring the quality and consistency of the data. The objective of our data cleaning was to ensure the dataset's integrity and quality for precise analysis. As a team, we aimed to identify and rectify inaccuracies and inconsistencies, handle missing values thoughtfully, and standardize data formats across the board. Our collective goal was to transform the raw data into a pristine, dependable resource supporting robust statistical analysis and effective machine learning modeling, setting the stage for deriving valuable insights.

**Handling Missing Values:**

Given the significance of certain columns like 'About the game' and 'Developers', where null values were present, appropriate measures were taken to address these gaps. The decision was made to drop the column for columns with negligible null values or those non-essential to the analysis (e.g., 'Score rank').



**Data Type Conversions:**

The 'Release date' column was converted from an object data type to a datetime format to facilitate time-series analysis. Similarly, numerical conversions were applied where necessary to ensure correct data type alignment for computational operations.

**Text Data Cleaning:**

Textual information within 'About the game', 'Supported languages', and categorization fields like 'Categories', 'Genres', and 'Tags' underwent cleaning. This process included standardizing text formats, parsing lists, and extracting relevant keywords to enable text analysis techniques like TF-IDF or NLTK.

**Feature Engineering:**

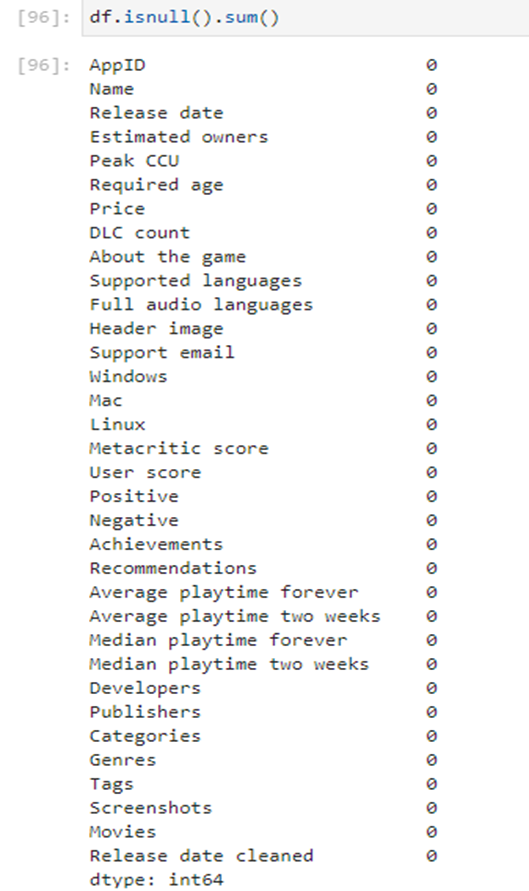
New features were derived from existing columns to enhance the dataset's analytical value. For example, create a binary feature from 'Price' to distinguish between free and paid games or aggregate user ratings into a composite score.

**Normalization:**

For fields with multiple entries like 'Supported languages' and 'Developers', normalization was performed to ensure a consistent structure. This enabled easier aggregation and analysis of the data.

**Final Dataset:**

The preprocessing steps resulted in a cleaned and enriched dataset, ready for exploratory data analysis and model building for game genre prediction.

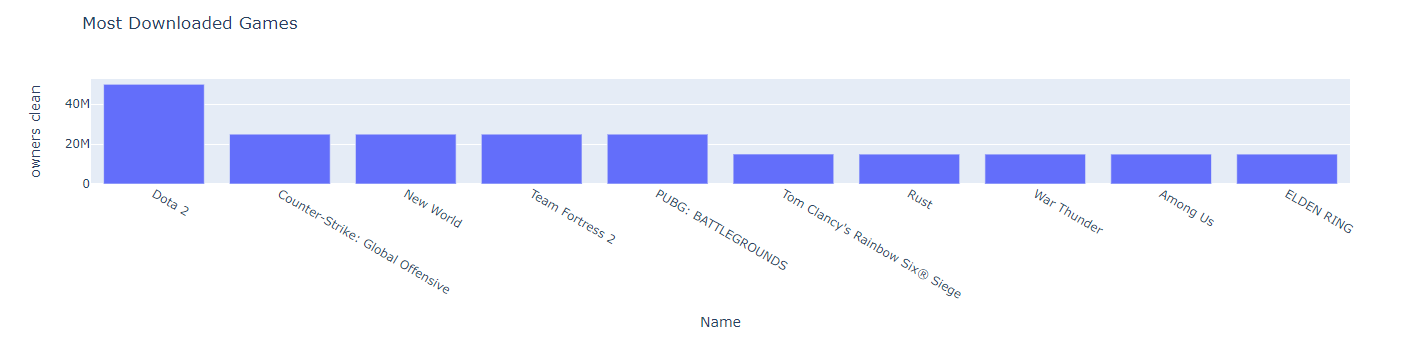


The result was a curated and analytically ready dataset primed for exploring game genres and user preferences on the Steam platform.

**Exploratory Data Analysis (EDA):** This data science project explores the dynamics of Steam game success through an in-depth analysis of genre, categories, and device platforms. The study uses exploratory data analysis to uncover patterns and trends that influence a game's reception in the market. The project seeks to provide valuable insights for stakeholders by employing EDA techniques to facilitate informed decision-making in game development, marketing, and platform optimization.

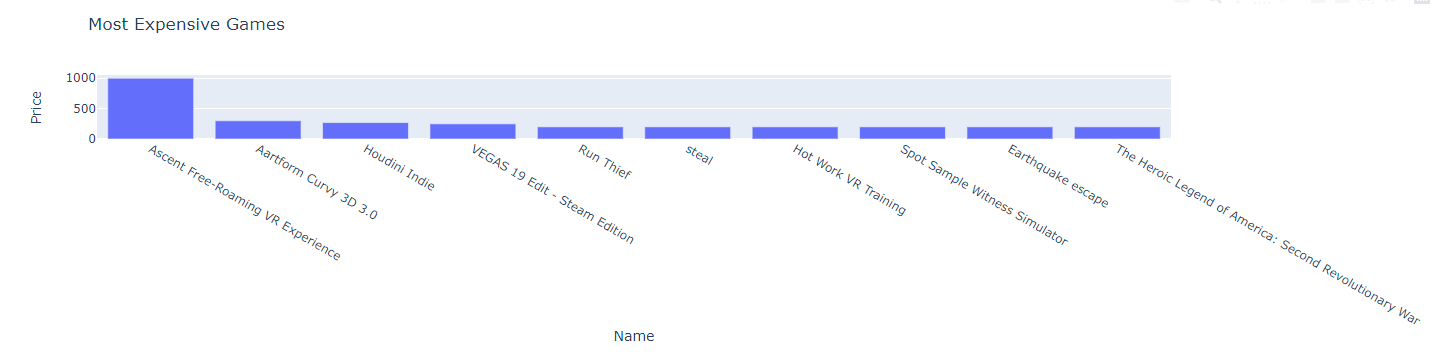
So, After the data collection and preprocessing stages were complete, our team transitioned to Exploratory Data Analysis (EDA) with clear objectives in mind. The goal of EDA was to uncover underlying patterns, spot anomalies, test hypotheses, and check assumptions with the help of summary statistics and graphical representations.

We aimed to better understand the data’s structure and the relationships between variables. This involved visualizing distributions, identifying trends, and exploring correlations within our dataset. EDA was pivotal for formulating informed assumptions for more complex analyses and ensuring that our subsequent modeling was grounded in a thorough initial data investigation.

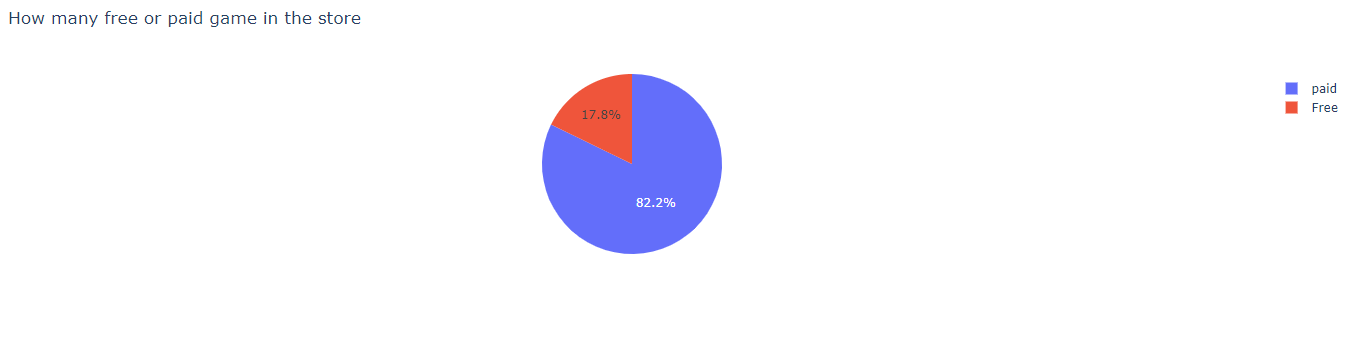


The bar chart above illustrates the "Most Downloaded Games" on Steam. Each bar represents the number of owners for a specific game. The horizontal axis lists the names of the games, while the vertical axis shows the number of owners, scaled in tens of millions.

From the chart, "Dota 2" appears to have the highest number of owners, followed by "Counter-Strike: Global Offensive", "New World", "Team Fortress 2", and others, indicating these games' popularity and widespread adoption among users of the platform.



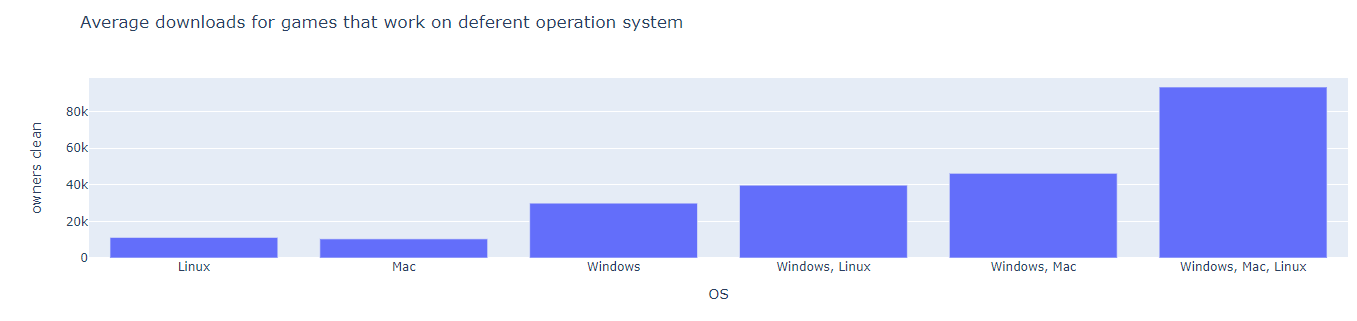
The bar chart above focuses on pricing, highlighting the "Most Expensive Games" in our collection. It reveals that "Ascent Free-Roaming VR Experience" commands the highest price, significantly more than other titles listed. This chart serves as a price comparison tool, revealing a wide range of pricing strategies employed in the market. The steep drop after the first title suggests that such high pricing is an outlier rather than the norm. Understanding this distribution helps assess how price points might relate to game features, exclusivity, and user willingness to pay.



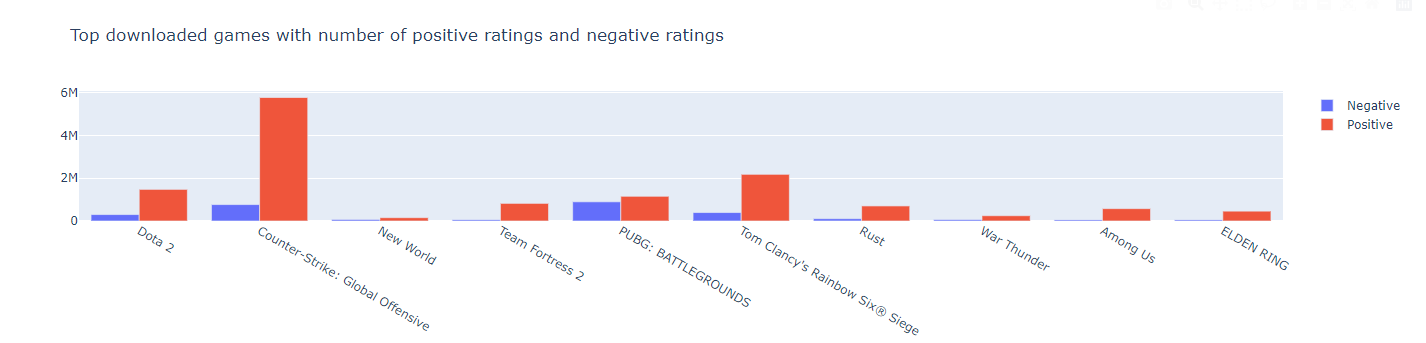
The pie chart shows that most games available in the store are paid, with approximately 82.2% requiring a purchase, while 17.8% are free to download and play. This indicates that the platform's catalog is largely driven by paid titles, which may reflect a business model focused on direct sales revenue.

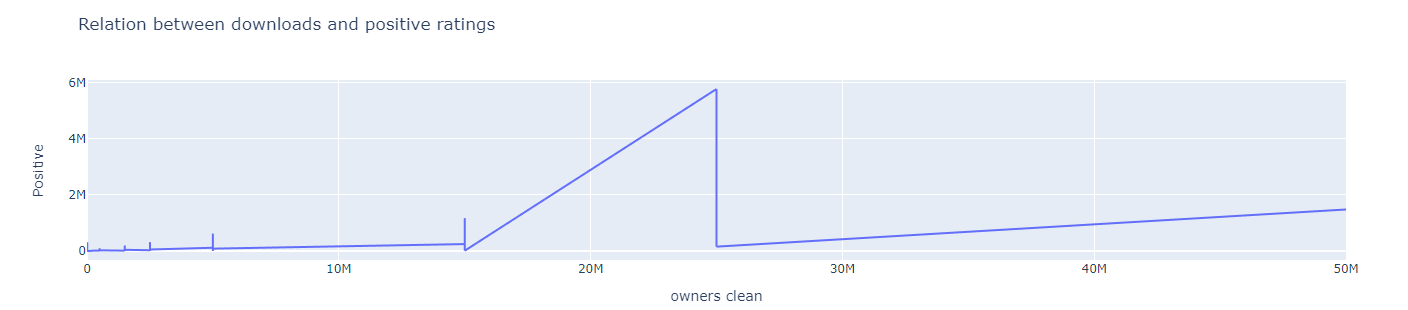
However, looking at the bar chart, we observe an interesting trend: free games have a higher average number of downloads (or owners) than paid games. This suggests that fewer free games are available; they tend to be downloaded more frequently, potentially due to the absence of a price barrier, leading to greater accessibility and wider distribution among users.

Incorporating both perspectives, the analysis suggests that while the platform's strategy might lean towards selling games as its primary revenue source, free games play a significant role in attracting users, possibly serving as a gateway for new users or as a means to keep the user base engaged with the platform. The higher average downloads for free games could also contribute to network effects, where an increased user base can attract more developers to the platform, create vibrant communities, and encourage in-game purchases or monetization through other avenues.

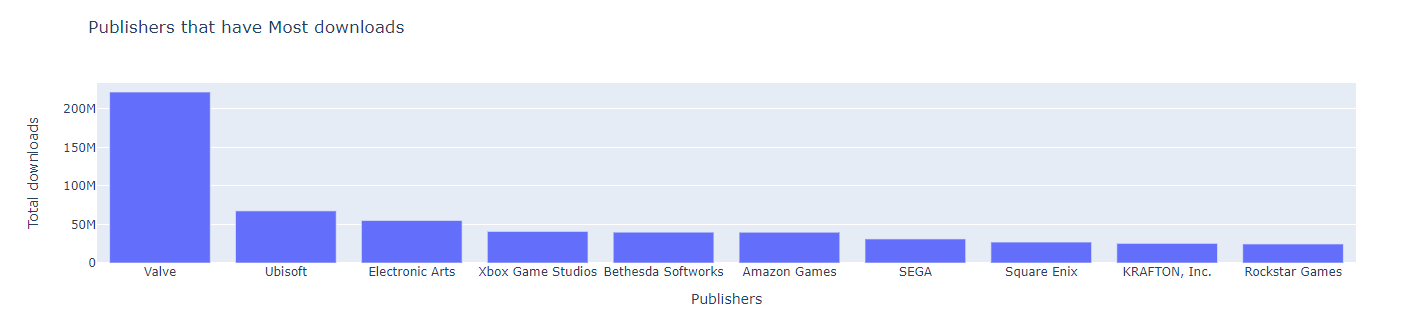


The bar chart titled "Average downloads for games that work on different operating systems" provides insight into the popularity of games across various platforms based on their operating system (OS) compatibility. The data suggests a correlation between the number of supported operating systems and the average downloads a game receives, underlining the importance of cross-platform compatibility in maximizing a game's potential user base and download numbers.

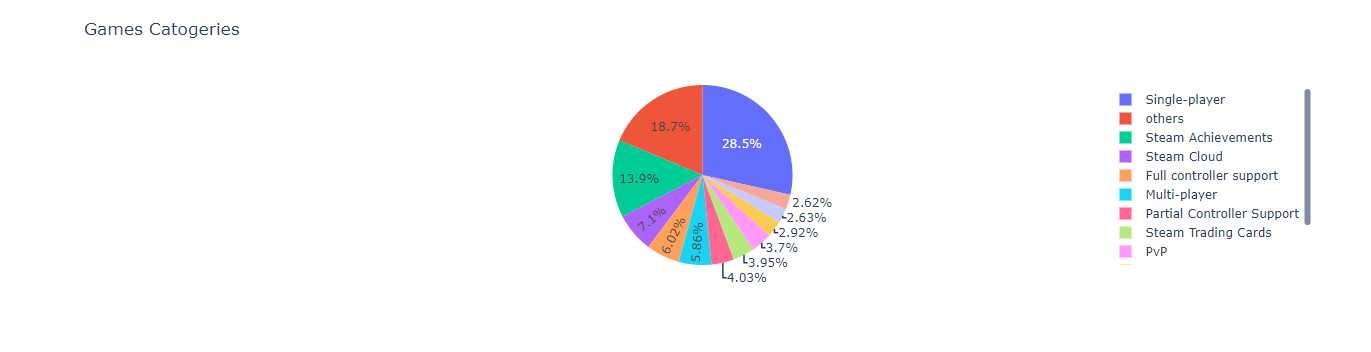




The bar chart and line graph collectively illustrate that top-downloaded games generally garner a significant number of positive ratings, with "Counter-Strike: Global Offensive" notable for its exceptionally high positive response. The data suggests that popularity, as measured by downloads, tends to coincide with favorable reviews, though negative ratings are also present, reflecting a spectrum of player reactions. This trend underlines the relationship between a game's reach and its reception, which can be a critical factor for developers and platforms in assessing their games' success and user satisfaction.



The "Publishers that have Most Downloads" bar chart depicts the total number of downloads associated with games published by various companies. Valve leads the chart by a significant margin, suggesting that games published by Valve are highly popular among users, potentially due to successful franchises or many titles available. Ubisoft, Electronic Arts, and other publishers like Xbox Game Studios and Bethesda Softworks also show substantial download figures but less than half of Valve's. Publishers like Amazon Games, SEGA, Square Enix, KRAFTON, Inc., and Rockstar Games follow with lower total downloads. This chart indicates market share in terms of game downloads among the leading publishers in the gaming industry.



The pie chart "Games Categories" displays the prevalence of different features in the game library, highlighting that single-player games are the most common, constituting over a quarter of the offerings. Steam Achievements and Steam Cloud features also have a significant presence, indicating the platform's emphasis on these value-added services. Other features like full controller support, multiplayer, and PvP are less common but still represent key categories. The substantial "Others" segment suggests a wide variety of additional features and genres within the game library. This distribution provides insights into player preferences and potential market opportunities for game development.

In the below image we can see a clear upward trend in the number of games over the years, with a particularly notable increase in recent years. For instance, in 2022, there were 13,961 games, which is higher than any previous year on the list. This could reflect the growing game development industry, the platform's expansion, or both. The number of games increased significantly starting from the early 2010s, which could correlate with the rise of digital distribution platforms that make it easier for developers to publish and distribute games.





The Action genre is the most popular among the three in terms of average downloads, followed by the Casual genre, with Indie games being the least downloaded on average. This could imply that Action games have a broader appeal or are more in demand among the genres presented.

**Feature Selection:**

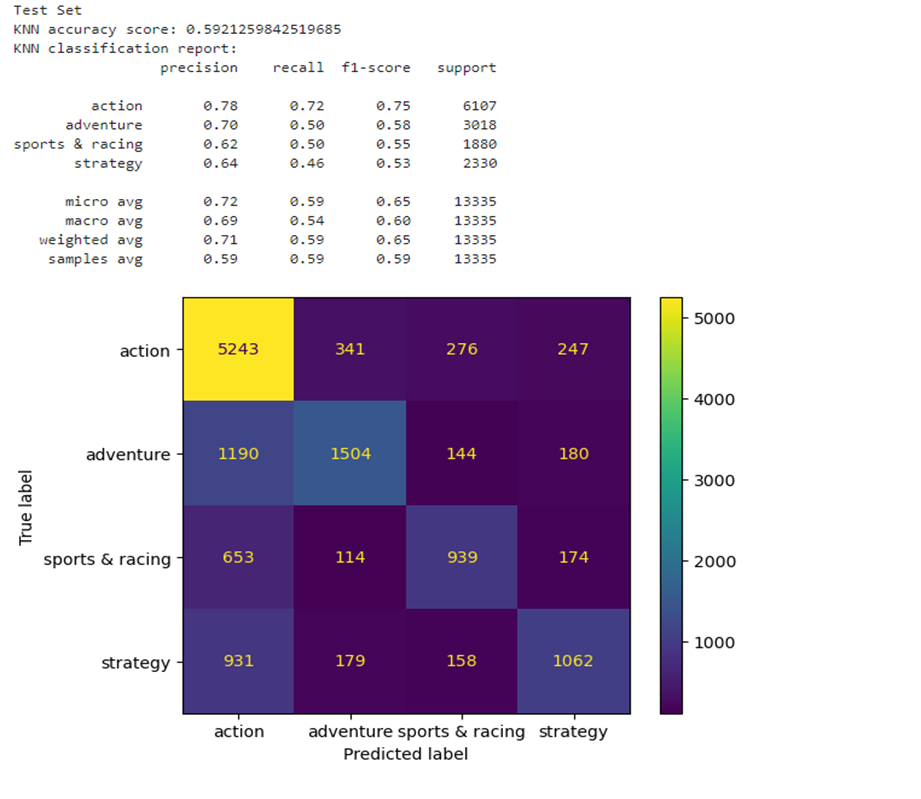
In our dataset of games, which comprises 39 different columns with various data types such as numerical, categorical, image, and text, we've identified that the 'genres' column is significantly influenced by the content of 'About the Game', 'Review', 'Categories', and 'Tags' columns. Due to its sparse data, we've decided to exclude the 'Review' column from our analysis. Furthermore, to streamline our computations, we've also chosen to omit the 'Categories' column since it contains information similar to that in the 'Tags' column. Therefore, we will use the 'About the Game' and 'Tags' columns as our features for training our model to classify game genres.

**Model Selection:** Model selection is crucial to enhance performance and predictive accuracy. In the Steam Game Dataset context, where tasks range from improving game genre classification to predicting game genres, it is essential to begin by establishing a baseline performance with various suitable ML algorithms. For this project, we have gone through multiple models and selected 5 models that we will be using in our project:

**1. K-Nearest Neighbor (KNN):**

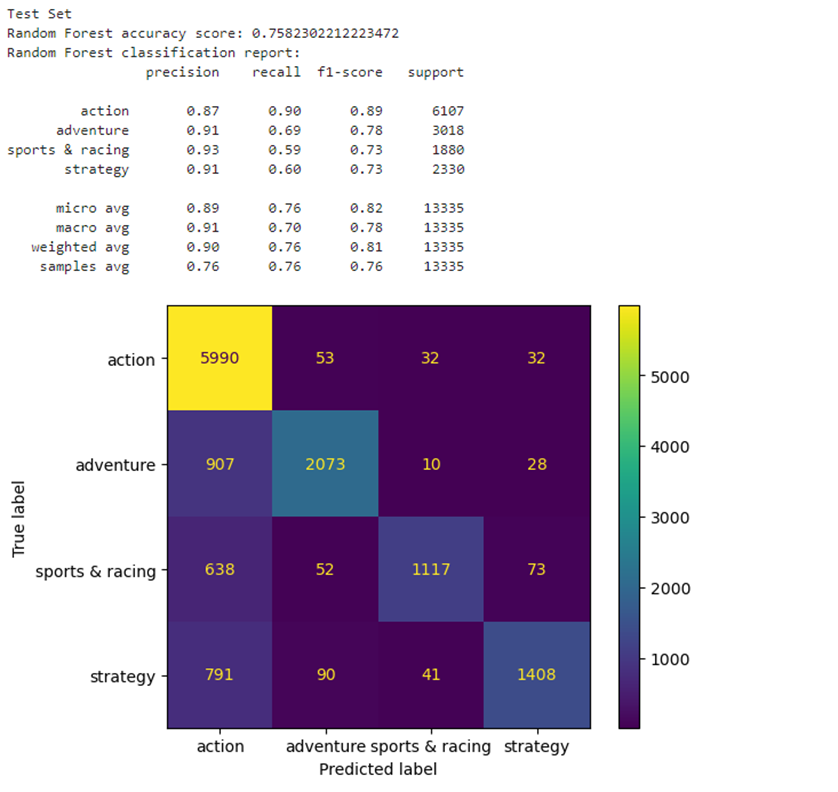
KNN is a simple, instance-based learning algorithm used for classification and regression. In both cases, the input consists of the k closest training examples in feature space. The output depends on whether KNN is used for classification or regression:

* In KNN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors.
* In KNN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors.



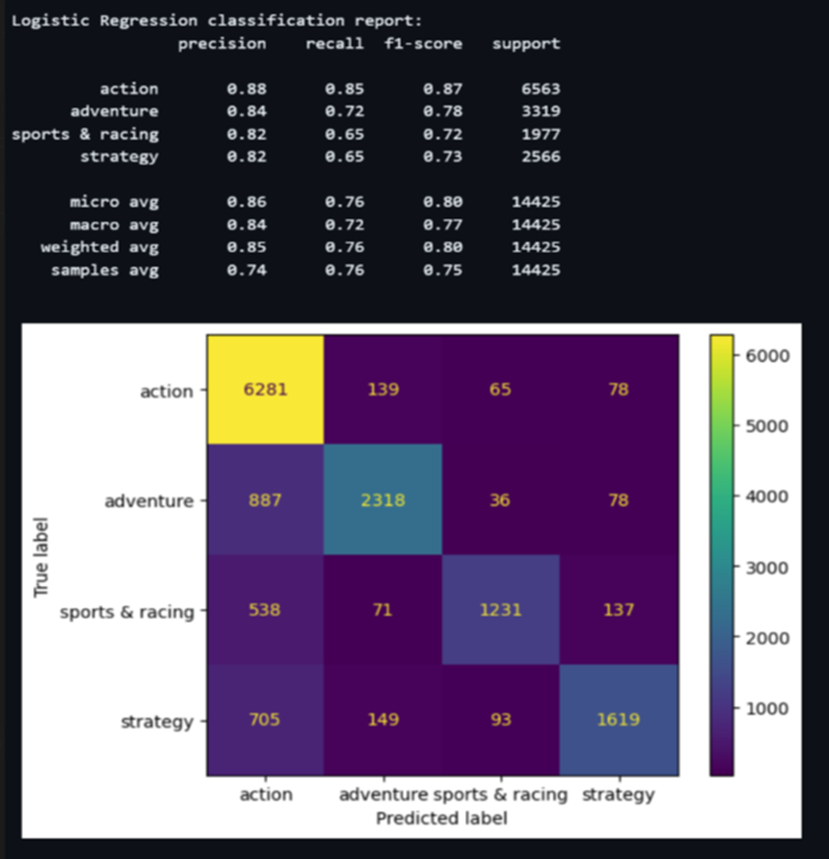
**2. Random Forest:**

Random Forest is an ensemble learning method for classification, regression, and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.



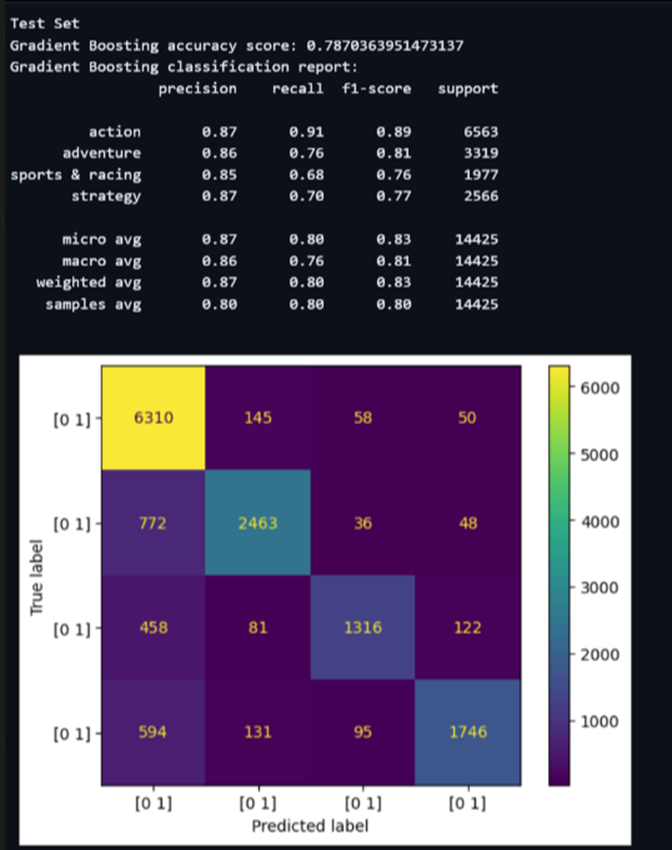
**3. One vs Rest (Logistic Regression):**

One-vs-rest (OvR) is a strategy that involves training a single classifier per class, with the samples of that class as positive samples and all other samples as negatives. This strategy is used for multi-class classification problems. Logistic Regression is a statistical model that, in its basic form, uses a logistic function to model a binary dependent variable, but in the context of OvR, it is adapted for multi-class classification.



**4. Gradient Boosting:**

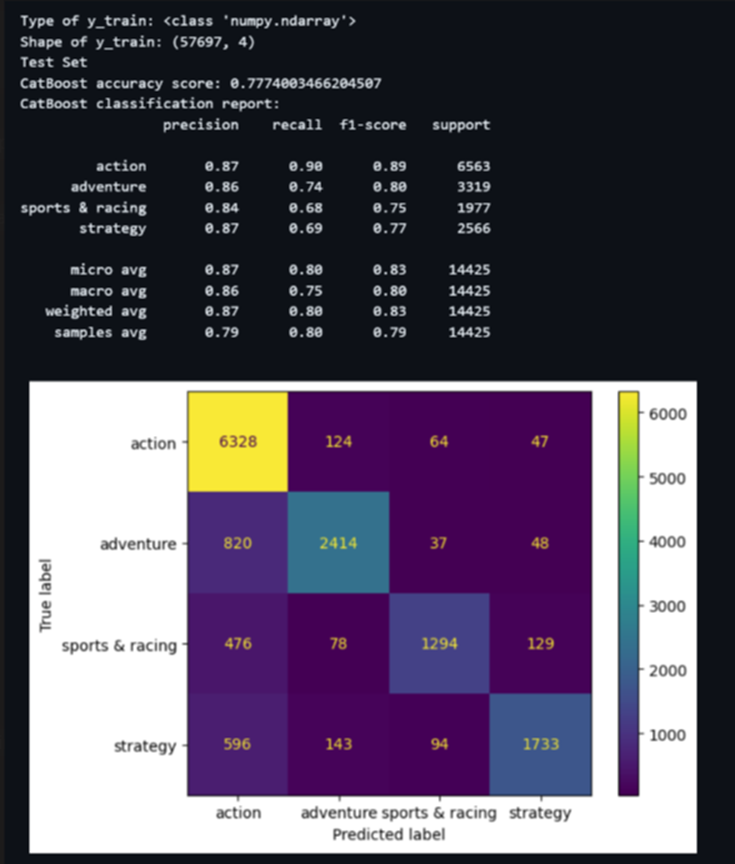
Gradient Boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion as other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.



**5. CatBoost:**

CatBoost (Categorical Boosting) is an algorithm for gradient boosting on decision trees. Yandex researchers and engineers developed it, which is especially powerful:

* It yields state-of-the-art results without extensive data training typically required by other machine learning methods.
* It provides powerful out-of-the-box support for categorical features (hence the name CatBoost for Categorical Boosting).



**Hyperparameter Tuning:**

We have implemented Grid Search for hyperparameter tuning of a Gradient Boosting model using the `GridSearchCV` class from the `sklearn.model\_selection` module.

The code defines a parameter grid for the gradient boosting model with the following parameters and respective ranges:

* `learning\_rate`: [0.1, 0.2, 0.3, 0.5]
* `max\_depth`: [5, 7, 9]
* `n\_estimators`: [100, 200, 300]

GridSearchCV is then instantiated with the gradient boosting model and the parameter grid. It is configured to perform a 3-fold cross-validation and to utilize all available CPU cores (`n\_jobs=-1`).

The output below the code snippet indicates that the best parameters found by GridSearchCV for the gradient boosting model are:

* `learning\_rate`: 0.1
* `max\_depth`: 9
* `n\_estimators`: 300

The best cross-validation score achieved with these parameters is 0.79.

Below the code output is a classification report for the test set showing precision, recall, and f1-score for different classes (action, adventure, sports & racing, strategy) along with the micro average, macro average, weighted average, and samples average.

The Gradient Boosting model's accuracy score on the test set is approximately 0.787.

**Model Evaluation:**

When evaluated on a test set, the Gradient Boosting Classifier shows approximately the highest 78.7 percent accuracy score.

The classification report provides detailed performance metrics, including precision, recall, and F1-score for each class (action, adventure, sports & racing, strategy) and aggregated metrics (micro average, macro average, weighted average, samples average).

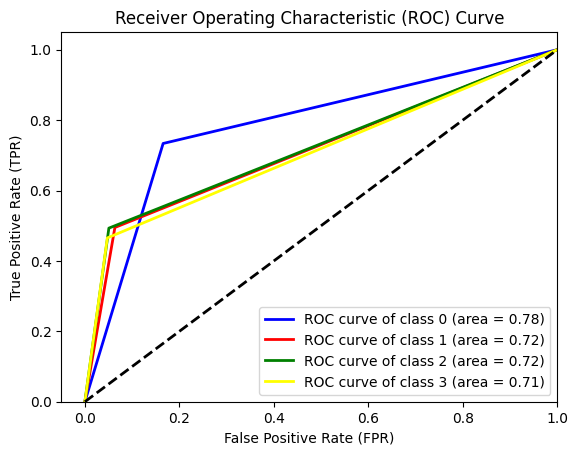
From the evaluation metrics, we can infer that:

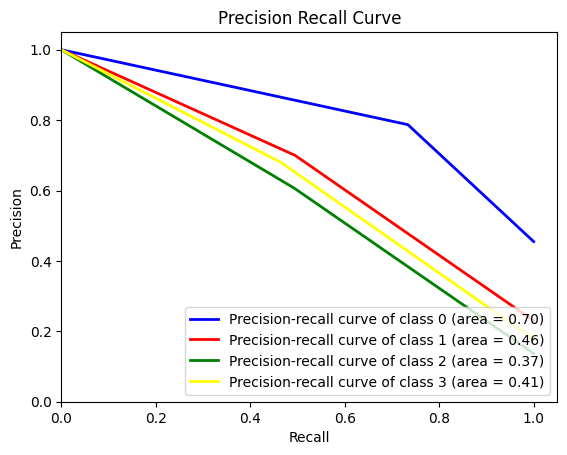
* The model performs well on the test set, with precision and recall above 0.85 for the 'action' genre, which seems to be the best-performing one.
* The 'strategy' genre has the lowest recall rate, suggesting it's more challenging for the model to identify correctly.
* The F1 scores are relatively high across all genres, indicating a balanced precision-recall trade-off.

Model evaluation typically includes these metrics and may involve analyzing confusion matrices, ROC curves, and precision-recall curves to fully understand the model's performance across different thresholds and classes. The provided metrics are useful for getting a general sense of how well the model is performing, especially if you compare it to a baseline model or to previous iterations of the model before hyperparameter tuning.

Here are the ROC curves and Precision-Recall curves for all our models-

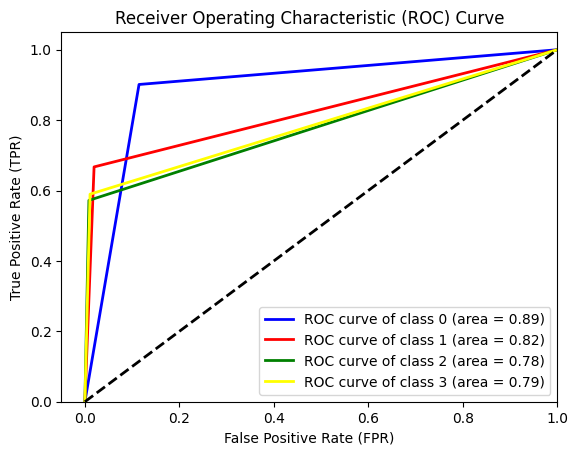
1. **KNN:**

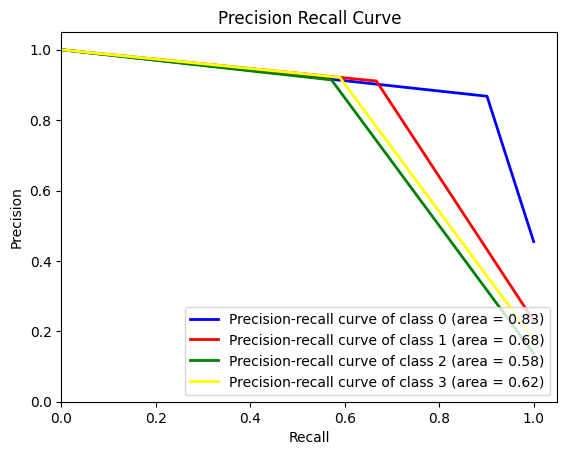
****

****

* The KNN classifier demonstrates reasonable performance across all classes, with class 0 being the most distinguishable.
* The AUC values from the ROC curve indicate that the classifier can distinguish between positive and negative classes across all categories.
* The PR curves reveal that while the classifier has a good balance of precision and recall for class 0, it struggles more with classes 1, 2, and 3, with class 2 being the most challenging for the model.
* The performance on class 0 is generally satisfactory, but there's room for improvement, especially for other classes, which might benefit from further model tuning or a more sophisticated algorithm.

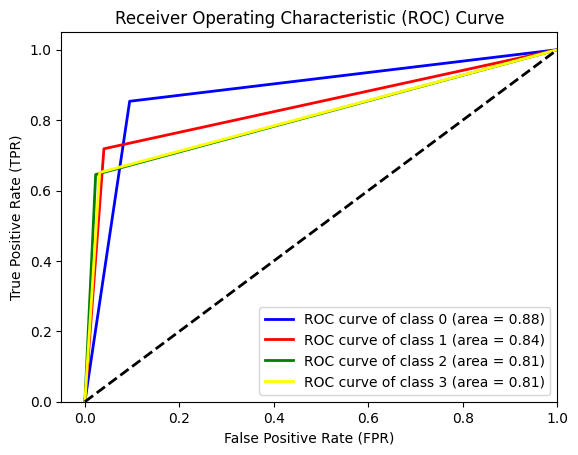
1. **Random Forest:**

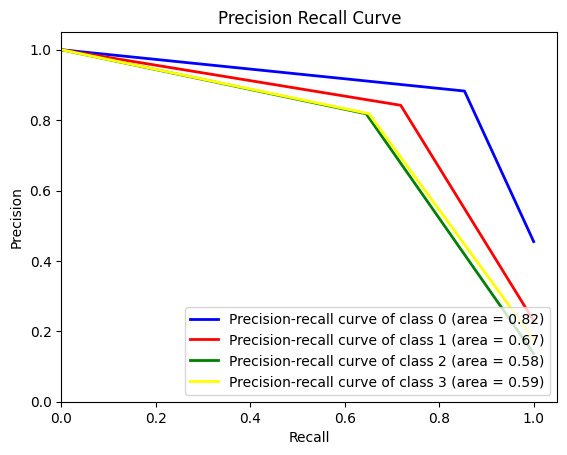
****

****

* The Random Forest classifier can distinguish between the classes, with class 0 being the most accurately predicted.
* The ROC curve suggests a good separation capacity across all classes, and the classifier appears to perform reasonably well in distinguishing positive from negative instances.
* The PR curves indicate that the classifier's precision is highest for class 0, but there is a notable drop-off for the other classes, especially class 2, which has the lowest precision-recall balance.
* These insights can inform potential areas of focus for model improvement, possibly through further hyperparameter tuning, feature engineering, or by addressing class imbalance if present in the dataset.

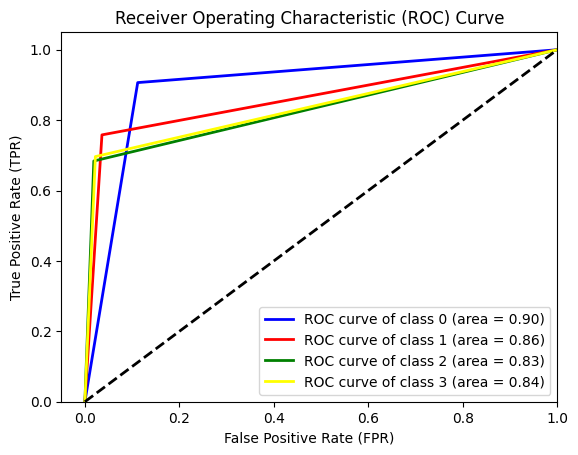
1. **One vs Rest(Logistic Regression):**

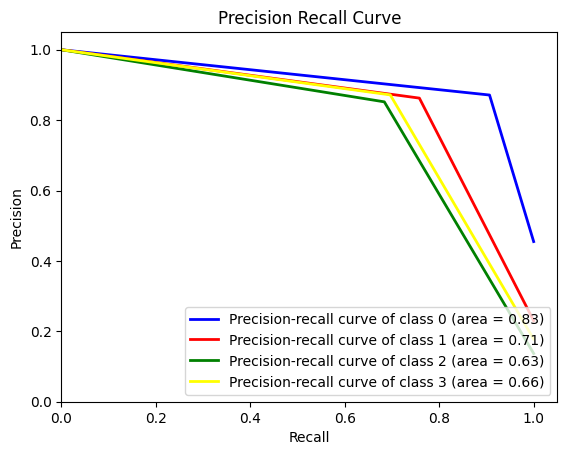
****

****

* The One-vs-Rest logistic regression classifier shows a strong discriminative ability for all classes as indicated by the ROC curves, with Class 0 being the best differentiated among them.
* The PR curves show that while the classifier is quite precise and sensitive for Class 0, its performance is less effective for the other classes, as indicated by lower precision-recall balance. This could be due to a number of factors, including class imbalance or less distinct class boundaries in the feature space for these classes.
* Combining ROC and PR curves provides a comprehensive view of the classifier's performance. While the ROC curve suggests good overall performance, the PR curve highlights areas where the classifier's performance could be improved, especially for Classes 1, 2, and 3.
* When dealing with imbalanced datasets, the PR curve is often more informative than the ROC curve, as it focuses on the classifier's performance with the positive class, which is typically the minority class in such scenarios.

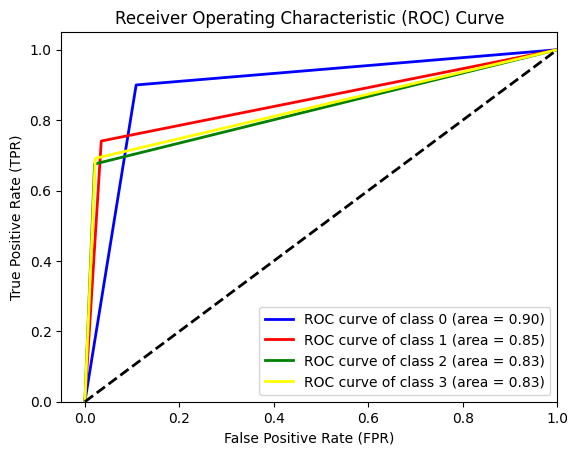
1. **Gradient Boosting:**

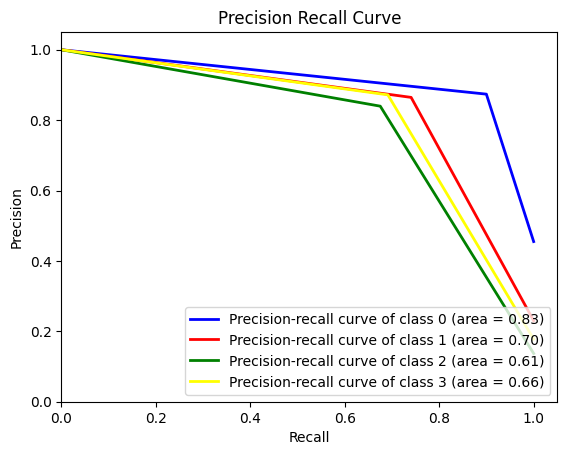
****

****

From these images, we can conclude that the Gradient Boosting model used to generate these curves performs well in classifying the different classes, with some variation among them. Class 0 is the easiest for the model to predict accurately, while Class 2 is the most challenging. The model has a good balance of precision and recall for Class 0, but the precision for the other classes, especially Class 2, is not as high.

1. **Cat Boosting:**

****

****

The CatBoost model shows strong performance distinguishing between the different classes, with Class 0 being the most distinct. The ROC curves indicate that the model balances false positives and true positives well. However, when it comes to precision and recall, there is some variation among the classes, with Class 0 performing best. This suggests that the CatBoost model is quite effective for this multi-class classification task, although the precision for predicting some classes (especially Classes 2 and 3) could potentially be improved.

**Conclusion:**

The machine learning model you've shown can be quite beneficial in real-life scenarios, particularly in the domain of video game analytics or recommendation systems. Here's how:

1. **Game Recommendation Systems:** The classifier could be part of a recommendation engine that suggests video games to users based on their preferences. For example, if someone likes games from the "action" genre, the model can predict and recommend new action games that the user is likely to enjoy.

2. **Market Analysis:** Publishers and developers can use such a model to understand current market trends. By analyzing which game genres score highly in terms of precision and recall, they can tailor their development to meet market demands.

3. **Inventory Management:** Retailers can use this classification model to manage inventory more efficiently by stocking more of the genres that are predicted to be popular or profitable, based on historical sales data and the model’s predictions.

4. **Personalized Marketing:** Marketing campaigns can be personalized based on the predicted preferences of different user segments. If the model predicts a user is likely to be interested in strategy games, then marketing content related to strategy games can be targeted to that user.

5. **Content Curation:** Online platforms like Steam or the PlayStation Store could use such models to curate content and feature games that are more likely to be popular within each genre category.

6. **Quality Control:** Developers could use the model to classify user reviews and feedback into genres, which can then be used to identify specific issues common to certain types of games and address them in future updates or releases.

7. **Time Series Forecasting:** By applying this model to historical data, businesses can predict future video game genre popularity trends, helping in strategic planning for development and marketing efforts.

8. **User Experience Enhancement:** For gaming platforms, this model could help improve the user interface by dynamically adjusting the visible content to match the user's predicted preferences, thus enhancing the overall user experience.

**References**

1. <https://www.researchgate.net/post/What_is_the_difference_between_the_three_Machine_Learning_models>
2. <https://www.kaggle.com/datasets/fronkongames/steam-games-dataset>
3. <https://www.statista.com/statistics/552623/number-games-released-steam>
4. [1.12. Multiclass and multioutput algorithms — scikit-learn 1.3.2 documentation](https://scikit-learn.org/stable/modules/multiclass.html)
5. <https://www.kaggle.com/datasets/luthfim/steam-reviews-dataset>