

PQC – Business Report

Table of Contents

- Executive Summary
- Business Understanding
 - Business Problem
 - BACCM Elements
- Data Understanding, Preparation and Exploration
 - Data Overview
 - Data Preparation
 - Data Exploration and Visualization
- Machine Learning Approach
 - Data Preparation
 - Feature Engineering and splitting the data
 - Model Selection
- Model and Performance Metrics
 - Final Model
 - Performance Metrics
- Business Solution and Recommendations
 - Focus on Game Engagement
 - Balance Game Complexity
 - Leverage User Interest Data
 - Optimize Player Configurations
 - Monitor and Improve Data Quality
 - Continuous Model Improvement

Executive Summary:

This report describes a machine learning approach for online gaming site Play Quest Conquer (PQC) to identify factors that influence ratings. The goal is to gain insights from the given data, build predictive models for relevant games, and provide optimal recommendations to improve user experience and performance. Years, age groups, and player profiles are tracked. Collaborative measures such as time between games and its relationship to viewership are also examined. Machine learning models are designed to predict average gameplay using factors such as game difficulty, popularity, and user satisfaction. Performance standards are evaluated and provide valuable information for the development of PQC and game support. Game measurement and user interaction. Additional information on data quality and correlation between variables is also provided.

1. Business Understanding:

1.1 Business Problem:

Play Quest Conquer (PQC) aims to understand the key factors that influence the rating of games on its platform. By analyzing these factors, PQC aims to adjust game development, acquisition, distribution, and promotion strategies to keep users happy and engaged.

1.2 BACCM Elements:

- Need: To increase user satisfaction and retention, factors affecting game ratings must be identified and understood.

- Stakeholders: Game Developers, Business Strategists, PQC's Market Research Team
- Value: Increase customer satisfaction with customized gaming products and marketing strategies.
- Solution: Machine learning model predict game rating based on game difficulty, popularity, user satisfaction and more.
- Context: In the gaming industry, especially on online platforms, user engagement and satisfaction are essential to business success,
- Change: Use data-drive game development and marketing strategies to increase user satisfaction and retention.

2. Data Understanding, Preparation, and Exploration

2.1 Data Overview: The dataset PQC_data.csv contains 17 columns which are as follows:

Column Name	Data type
Game_ID	int64
Game_Name	object
Released_Year	int64
Game_Type	object
Age_Category	object
Min_Players	int64
Max_Players	int64
Average_Complexity	float64
Complexity_Raters	int64
Average_Play_Time	int64
Owner_Number	int64
Trader_Number	int64
HighInterest_Number	int64
Interest_Number	int64
Rater_Number	int64
Comment_Number	int64
Average_Rating	float64

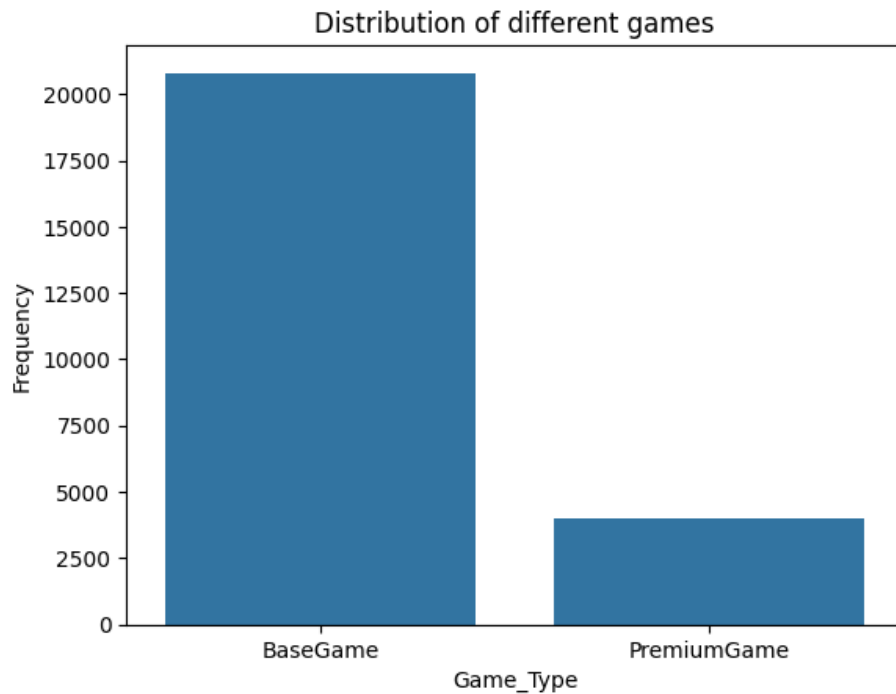
2.2 Data Preparation:

The data went some cleaning and preprocessing as outliers were there. The Game_Name has 7 missing values, since Game_name is not our important feature, we replace all the missing values with the name '**Unknown**'. The Released_Year column has some values as **99**, so it might be possible that there is some error, for correcting it we replace those values with the **median** Release Year. Some of the values in the Min_Player and Max_Player columns are **0**, it has possibly occurred due to some error, so for correcting it replace those values with the **median** player value.

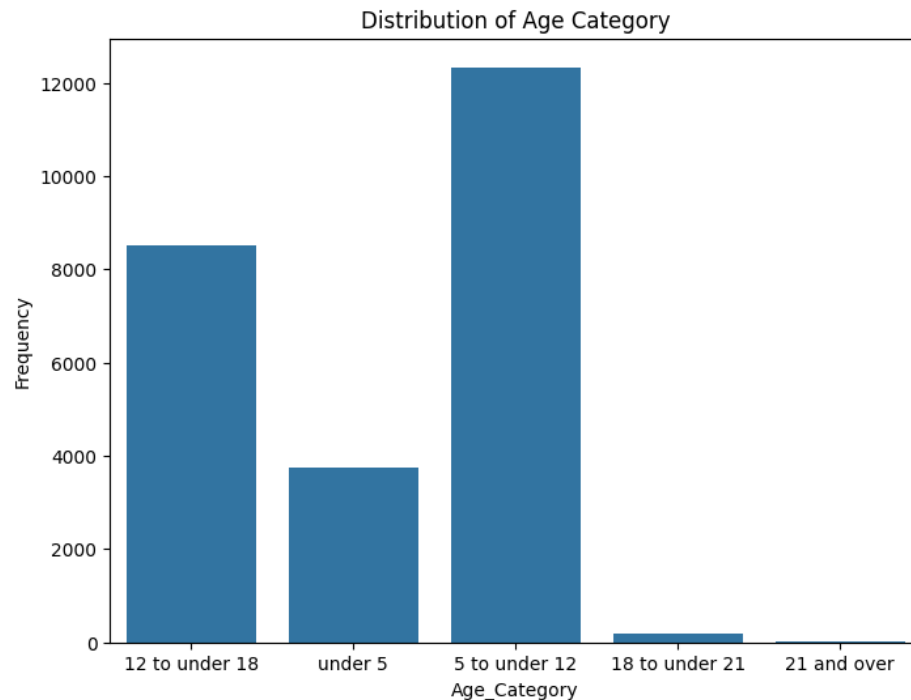
2.3 Data Exploration and Visualization:

General Information and game configurations:

1. Distribution of different types of games: There are two Game types, Base Game (20,796 games) and Premium Game(4017 games).



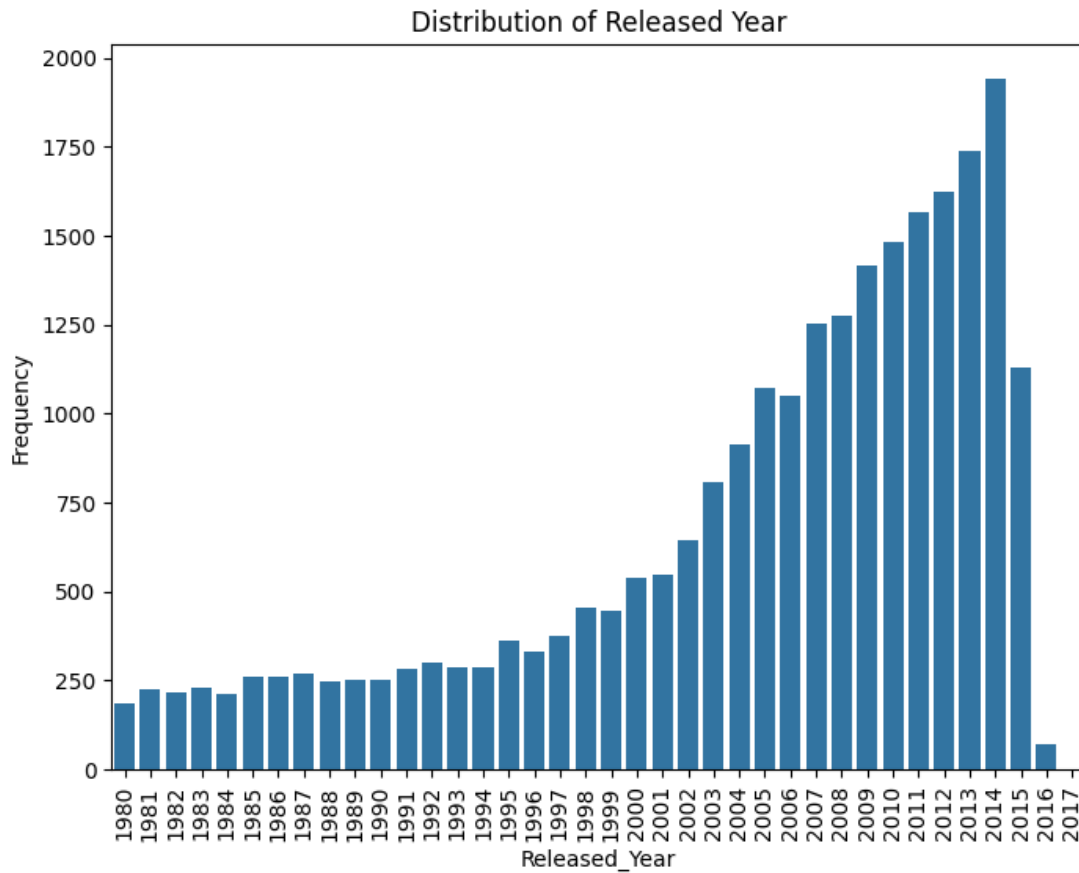
2. Distribution of Age Category: There are total 5 different age categories namely, 'under 5'(12335 games) , '5 to under 12'(8520 games) , '12 to under 18'(3751 games) , '18 to under 21'(183 games) and '21 and over'(24 games).



3. Year of Release: Release year ranges from 1980 to 2017. It is shown that there is a significant increase in the number of releases around the early 2000s.

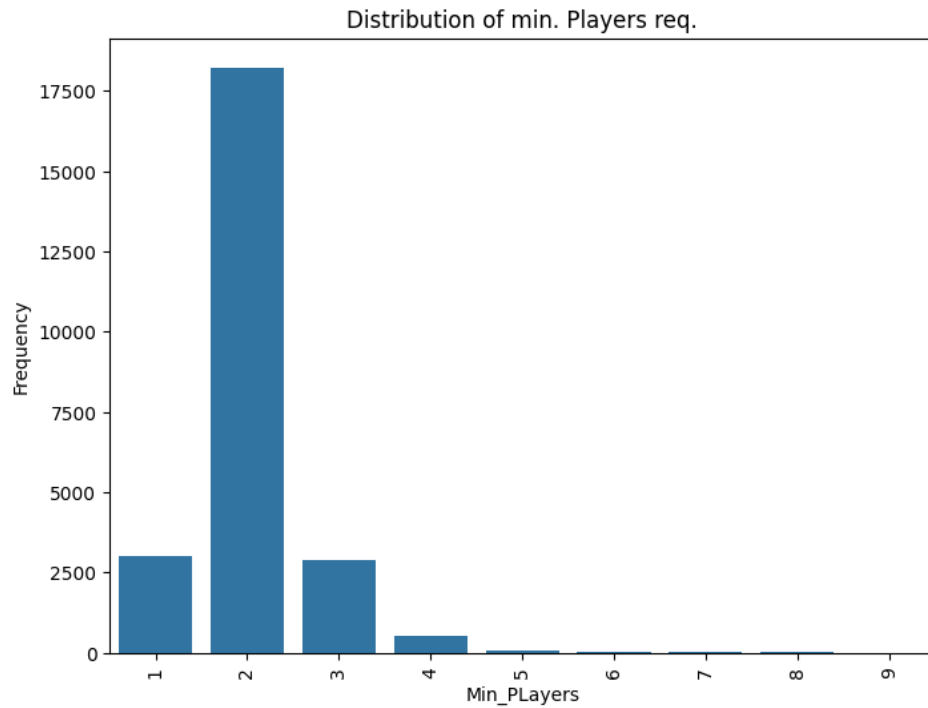
Release Year	No. of Games
1980	185

1990	254
2000	537
2010	1484
2015	1129
2016	73
2017	2



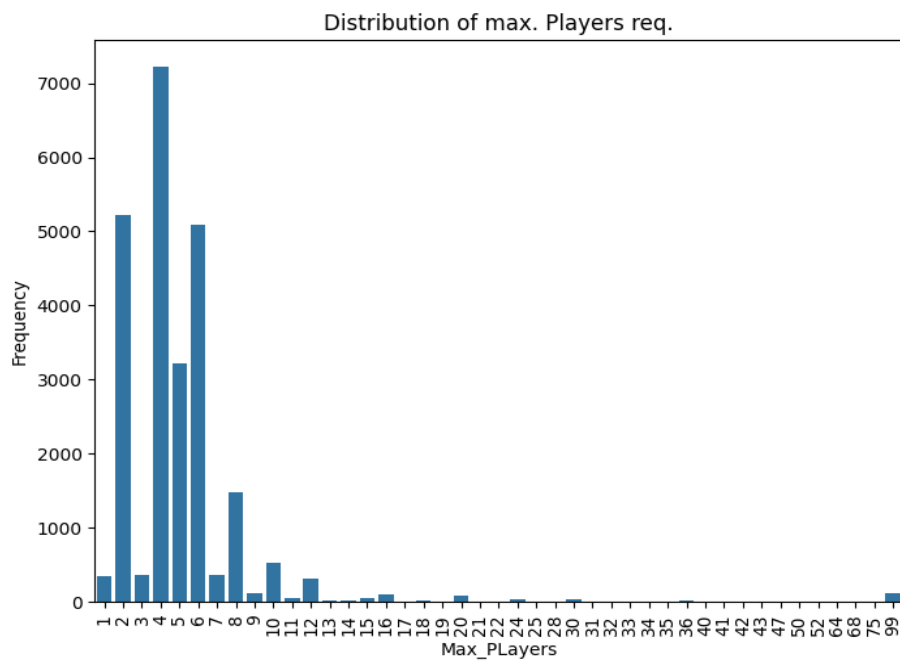
4. Distribution of Minimum Number of Players Required:

Minimum Number of Players	No. of games
1	3011
2	18177
3	2876
4	532



5. Distribution of Maximum Number of Players required:

Maximum No. of Players	No. of games
2	5225
4	7186
5	3215
6	5083
8	1485
99	108

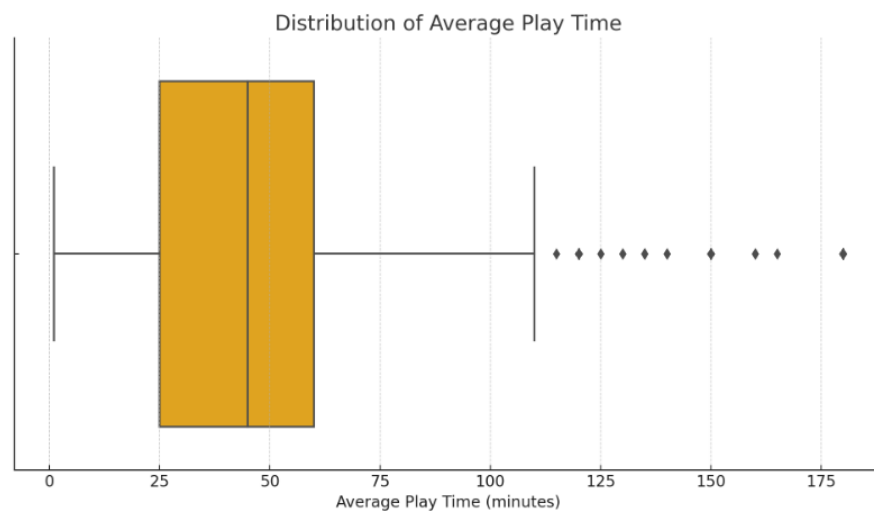
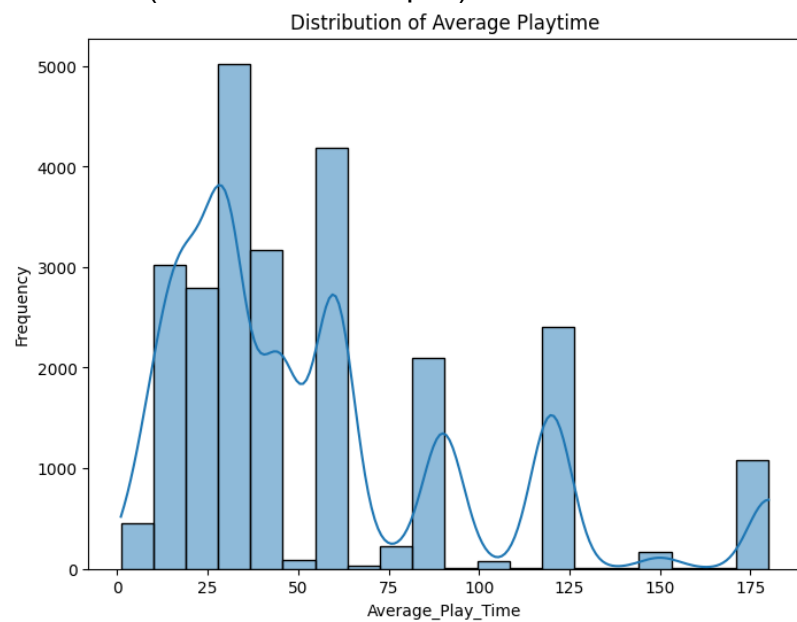


Game Engagement:

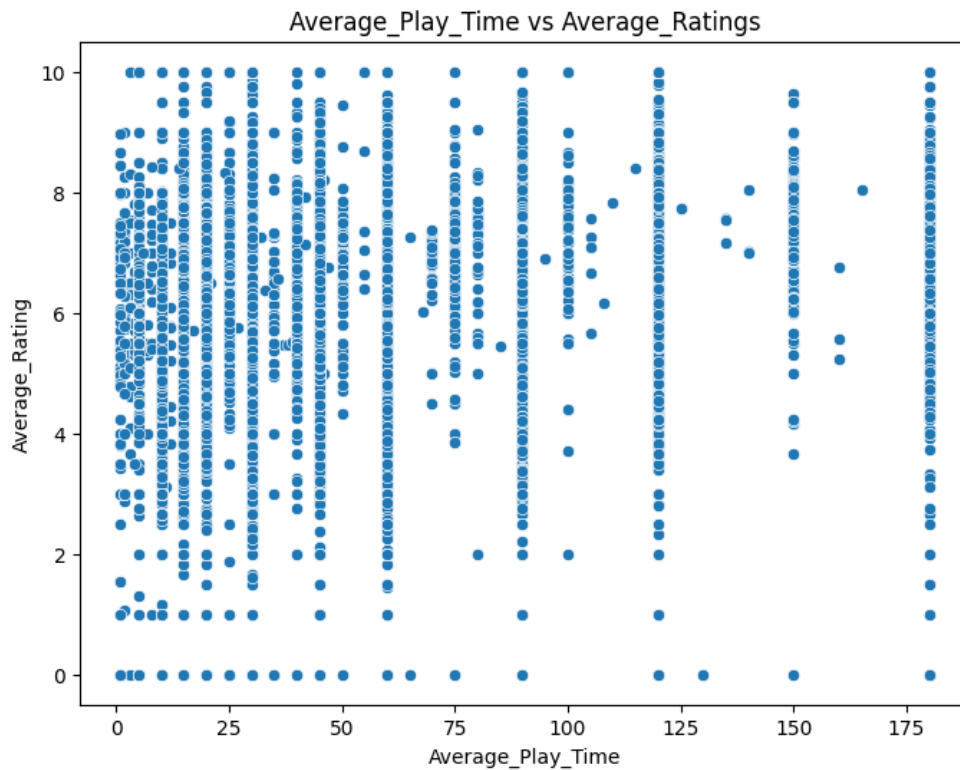
1. Distribution of Average Playtime:

Central Tendency	minutes
Mean	55.21
Standard Deviation	42.59
Minimum	1
25 th percentile	25
Median (50 th percentile)	45
75 th percentile	60
Maximum:	180

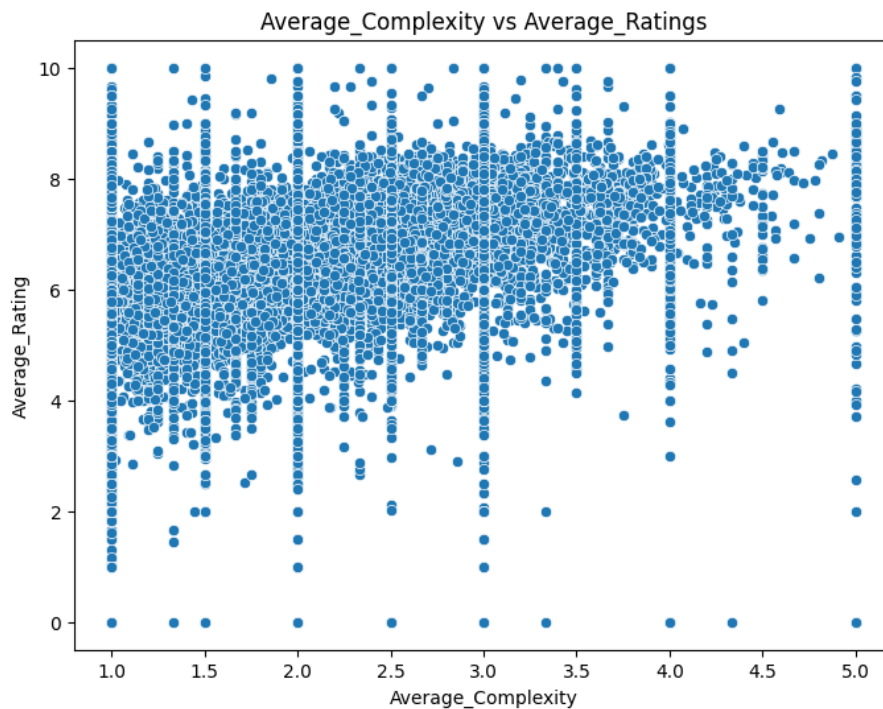
Outliers Identified: 3672 (shown in the box plot).



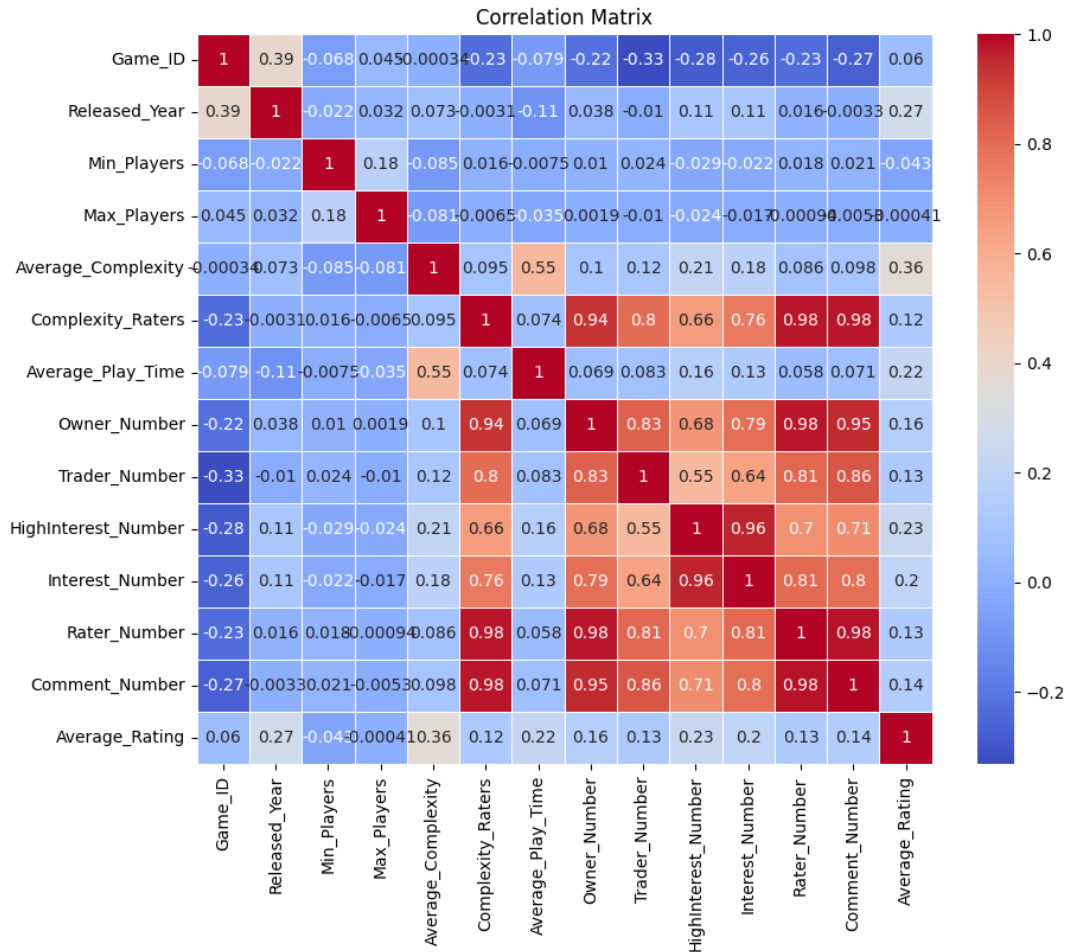
2. Correlation between Playing time and average ratings: The correlation coefficient between Playing time and average ratings is 0.217 which indicates a loose positive linear relationship.



3. Correlation between Game Complexity and Average Ratings: The correlation coefficient between Average Complexity and Average Ratings is 0.361, which signifies a better positive linear relationship than in between playing time and ratings.



4. Correlation between Game Configuration/Popularity/Interest and Average Ratings: The correlation coefficients lie between -1 to 1 where -1 represents a perfect negative linear relationship and +1 represents a perfect positive linear relationship and 0 represents there is no linear relationship.



3. Machine Learning Approach:

Data Preparation

Data layer has undergone a lot of cleaning and pre-processing. Missing value in Game_Name field has been changed to "Unknown". Fixed incorrect values in Released year, Min_Players and Max_Players fields using median values.

Feature Engineering and splitting the data

Once all the features provided in the dataset are used, no further feature engineering is required as the provided features are considered sufficient for analysis. The system uses an 80-20 split to measure performance standards. Cross validation was used to ensure consistency in model performance.

Model Selection

Use supervised learning to test a variety of algorithms: linear regression, tree pruning, random forests, and gradient boosting. The final model was selected based on performance evaluation, specifically mean square error (MSE) and R-squared (R^2) scores.

4. Model and Performance Metrics:

Final Model

The selected model is Random Forest which shows stronger prediction for average game compared to other models.

Performance Metrics

- Linear Regression: MSE = 2.0613 and $R^2 = 0.1824$
- Decision Tree: MSE = 1.5086 and $R^2 = 0.4017$
- Random Forest: MSE = 0.7387 and $R^2 = 0.7070$
- Gradient Boosting: MSE = 2.07613 and $R^2 = 0.6980$

The Random Forest model showed the best performance with an MSE of 0.7387 and a R^2 of 0.7070, indicating it is reliable for predicting game ratings based on the available features.

5. Discussion of Pros and Cons:

Pros

High accuracy: Random Forest model shows accuracy in prediction.

Interpretation: Results are interpretable and provide clarity to business decisions.

Cons

Potential overfitting: Due to complexity of random forest models, there is a risk of overfitting.

Limited generalizability: Due to the synthetic nature of the data set, the generalizability of the model will be limited.

Data biases: Data biases can affect predictive models and therefore should be continually monitored and updated.

6. Business Solution and Recommendations:

- Focus on Game Engagement

Findings:

Our analysis showed that there is a positive correlation between average play time and game ratings. Games with longer average play time will have higher ratings. This suggests that users who spend a lot of time in the game are more likely to rate the game well, indicating that engagement is an important factor in user satisfaction.

Recommendations:

PQC should prioritize purchasing and promotional activities that encourage longer-term engagement. There are several strategies to achieve this:

- In-game incentives: Offer features that reward users for playing longer, such as unlocking new levels, characters, or items when people spend a lot of time in the game.
- Engaging content: Create and promote games with rich, compelling stories and great gameplay that keep users engaged for a long time.

PQC can increase user satisfaction by focusing on games that typically generate longer playtimes, which can result in higher ratings and retention rates.

- Balance Game Complexity

Findings:

There is a positive relationship between the difficulty of the game and the ratings. However, this relationship is very subtle. Although the game's difficulty can be mea

sured by experienced players, extremely difficult games may discourage casual users, resulting in lower scores.

Recommendations:

As a game for both casual and hardcore gamers, PQC should develop and market games with the same balance:

- Easy to play yet challenging: Games should be designed to be accessible to new players or player-friendly while providing depth and challenge for those who are knowledgeable as gamers.
- Difficulty: Use adjustable difficulty levels that allow players to choose their preferred difficulty level.

Difficulty balance makes the game appeal to a wider audience, increasing overall user satisfaction and ratings.

- Leverage User Interest Data

Findings:

The number of users marking a game as "Ready" or "Modified" was positively correlated with the rating, suggesting that games with initial traction are more likely to receive positive reviews.

Recommendation:

PQC should use the "Happiness" and "Change" indicators to indicate game development and promotion ideas:

- Importance of development Editing: Improve and enhance games that yield high result first. These games have more potential for success and good reviews.
- Feedback from users whose games are marked as "Hot" or "Warm" for improvement and enhancement of these games.

By focusing on engaging games, PQC can improve customer satisfaction and ratings, leading to better business performance.

- Optimize Player Configurations

Findings:

Games that allow for a variable number of players will be rated higher. This change can improve social interaction, which is one of the key factors of user engagement.

Recommendations:

PQC should encourage and develop games with competitive options to accommodate different team sizes:

- Multiplayer Options: Design player configurations that allow flexibility in gameplay, allowing for solo, small groups or large groups.
- Foster a sense of community and encourage players to communicate with each other.

Optimizing player profiles and improving relationships can lead to higher engagements, better ratings, and more users.

- Monitor and Improve Data Quality

Findings: Several data quality issues were found during the analysis, such as missing values and inconsistent publication years. Bad information can lead to misunderstandings and poor performance models.

Recommendation:

To achieve better understanding and more effective models, PQC must establish a rigorous and valid data collection process:

- Data collection model: Use a standardized data collection set to reduce errors and inconsistencies.
- Review: Data sets are reviewed regularly to identify and correct missing values, inconsistencies, and inaccuracies.

Maintaining good data is crucial to making informed business decisions.

- Continuous Model Improvement:

Finding:

Machine learning models are good but can be improved with more data or additional features. Continuous improvement is necessary to maintain a correct and accurate model.

Recommendation:

PQC should focus on continuing to improve its predictive models by incorporating new data and testing advanced methods:

- New Data: Integrating other data, such as user feedback, gameplay data, and marketing data, to improve the accuracy of predictive models.
- Advanced algorithms: Try different machine learning algorithms and techniques (like deep learning) to improve prediction.

By constantly improving its models, PQC can maintain a competitive advantage in predicting the game and making data-driven decisions.

By focusing on game engagement, measuring vulnerability, using user satisfaction data, optimizing user data, improving data quality, and continuing to improve predictive models, PQC can increase user satisfaction and achieve better business results. These strategies will help PQC maintain its leadership position in the competitive online gaming industry and achieve long-term success.