ANALYZING PLAYER FEEDBACK IN STEAM REVIEW
ACROSS GAME GENRES

SAFIRA NURUL IZZA

UNIVERSITI TEKNOLOGI MALAYSIA

Table of Content

CHAPTER 4 RESEARCH DESIGN AND IMPLEMENTATION

CHAPTER 4

RESEARCH DESIGN AND IMPLEMENTATION

4.1 Introduction

This chapter contains graphical analysis of player review datasets from 5 genres with 25 games that have been collected, where insights from visualization and descriptive statistics will be revealed. Feature engineering will also be discussed in order to create, modify, and select features or variables in datasets to improve machine learning performance. This chapter also carries out feature extraction to convert textual data (reviews) into numeric, so that reviews can be combined with other numeric features (e.g., review length, play hours). Finally, a machine learning model will be introduced to generate initial insights.

4.2 Exploratory Data Analysis (EDA)

EDA is a critical step in understanding the structure, trends, and insights hidden in the data. For this project, graphical representations were used to visualize relationships within the data and the distribution of variables based on the feature engineering established in the methodology chapter.

4.2.1 Visualization and Descriptive Statistic

The goal of visualization and descriptive statistics is to simplify large datasets into easy-to-interpret graphical formats. This approach also helps present findings more effectively to stakeholders or non-technical audiences. Several types of charts have been chosen to present visualizations to help uncover hidden patterns, trends, and relationships in the data.

4.2.1.1 Sentiment Distribution

A pie chart was used to illustrate the sentiment distribution of the “Thumb Text” column, which contains the Recommended or Not Recommended values for each genre. This analysis helps understand the proportion of positive and negative reviews in each genre.

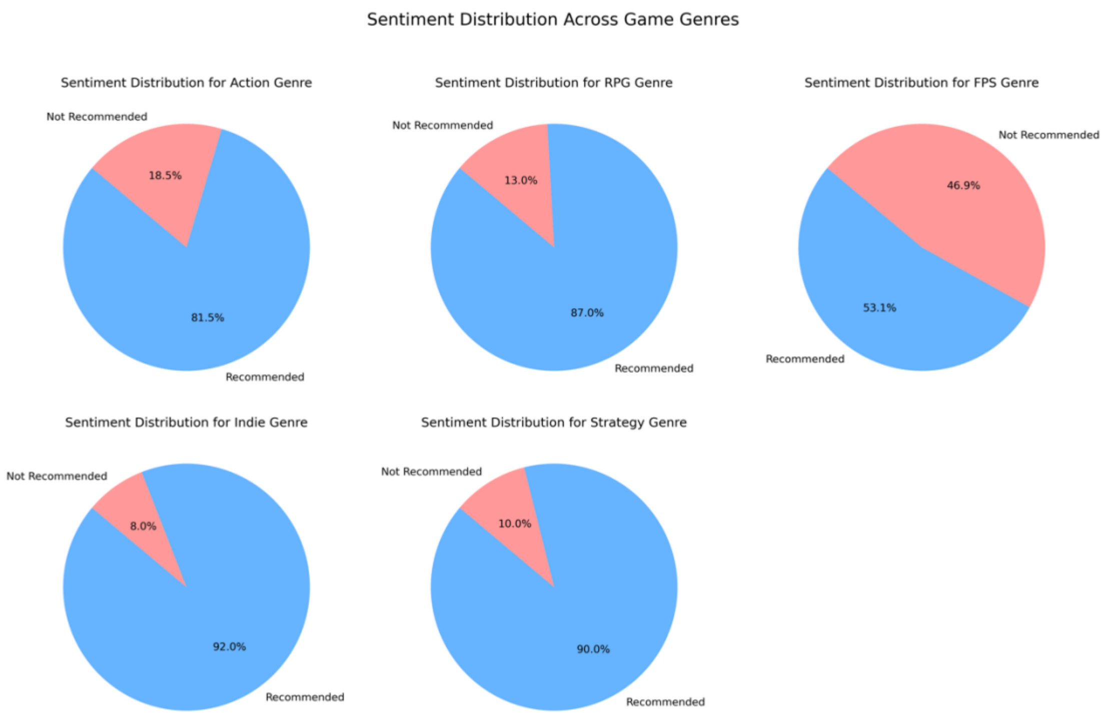

Figure 4.5: Sentiment Distribution Across Game Genres (Pie Chart)

Based on the pie chart, FPS or First-Person Shooter has the most negative feedback with a percentage of 46.9%, while Indie has the most positive feedback with 92% positive reviews.

4.2.1.2 Review Length Distribution

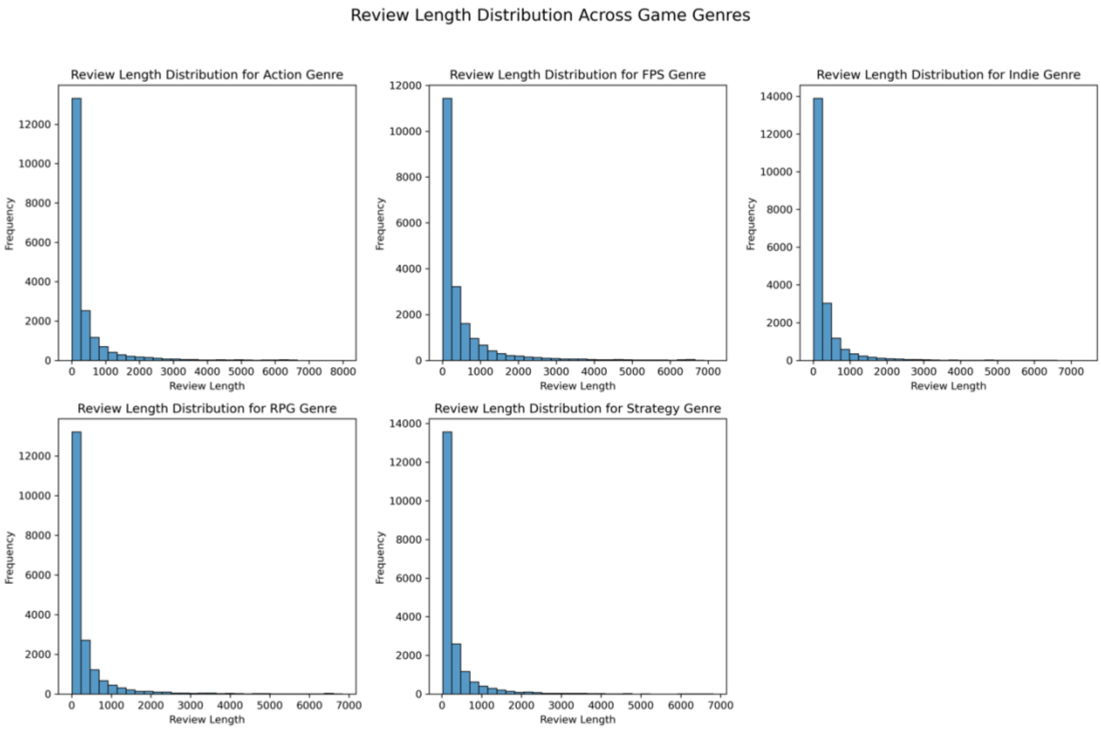

Figure 4.6: Review Length Distribution by Genre

The histogram chart depicts the distribution of review lengths, which shows patterns in the amount of player feedback. The distribution highlights that across all genres, most players leave short, concise reviews. Conversely, the longer the review, the fewer players leave such a review.

4.2.1.3 Review Trends Over Time

A line chart visualized the number of reviews posted per month/year for each genre.

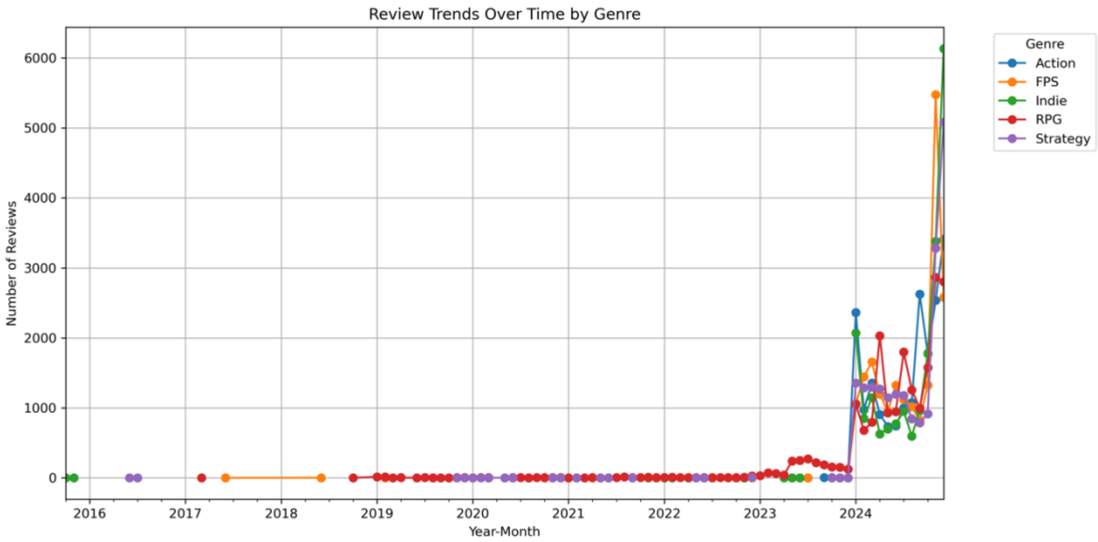

Figure 4.7: Review Trends Over Time by Genre

Key observations included:

1. Action: Reviews started from June 2021 to December 2024, with the highest review count in December 2024.

2. FPS: Reviews ranged from June 2017 to December 2024, peaking in November 2024.

3. Indie: Reviews spanned from October 2015 to December 2024, with a peak in December 2024.

4. RPG: Reviews were posted from March 2017 to December 2024, with the highest in November 2024.

5. Strategy: Reviews ranged from June 2016 to December 2024, with a peak in December 2024.

These trends highlighted periods of increased player activity, often corresponding to game updates or promotional events.

4.2.1.4 Play Hours Analysis

A box plot showed the distribution of play hours for each genre, providing insights into player engagement.

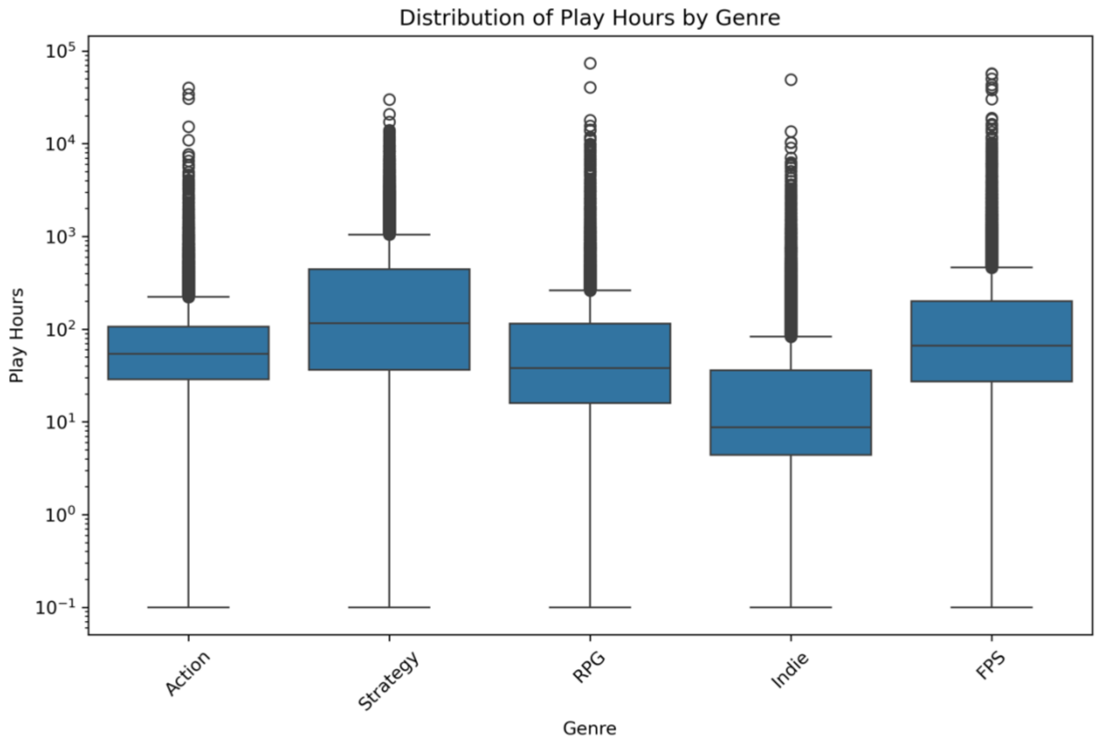

Figure 4.8: Distribution of Play Hours by Genre

Summary statistics included:

1. Action: Mean = 121.62, Median = 54.30
2. FPS: Mean = 272.45, Median = 66.10
3. Indie: Mean = 79.29, Median = 8.70
4. RPG: Mean = 137.79, Median = 37.80

5. Strategy: Mean = 486.52, Median = 116.00

The Strategy genre had the highest average play hours, indicating deeper player engagement, while Indie games had the shortest play times on average.

4.2.1.5 Censored Text

A pie chart was used to visualize the percentage of reviews containing censored text for each genre. From this chart, it can be seen in figure 4.9 that some inappropriate words were used by players in reviews about the game.

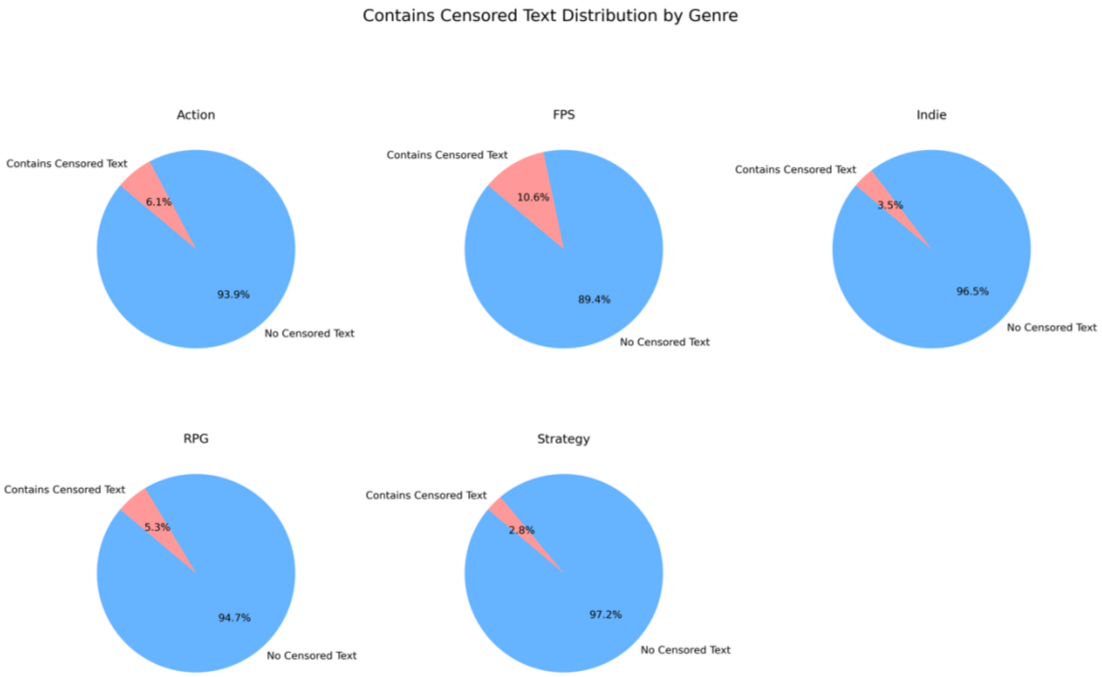

Figure 4.9: Distribution of Reviews Containing Censored Words Across Genres

4.2.1.6 Product Refunded

A pie chart was used to show the proportion of refunded and non-refunded reviews for each genre. From this chart, patterns of players requesting refunds for certain games can be observed, which may indicate the level of player satisfaction or dissatisfaction with the games.

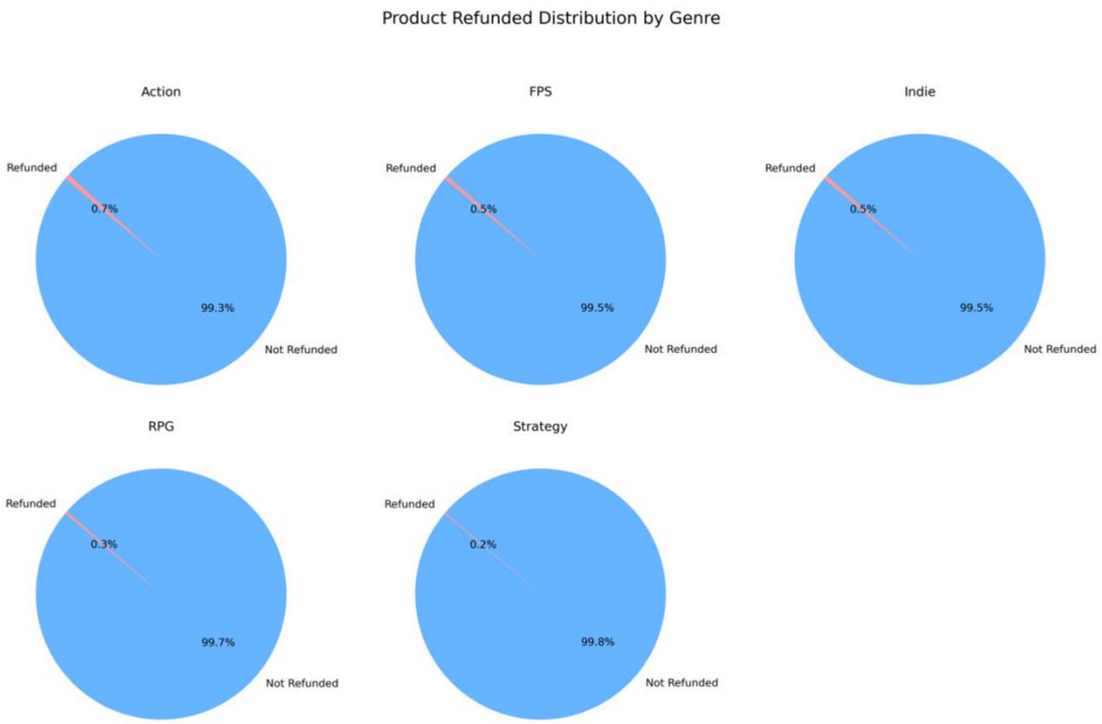

Figure 4.9: Distribution of Refunded and Non-Refunded Reviews Across Genres

4.2.2 Initial Insight Gained

Four different charts were created to generate graphical information from the datasets of the five genres, with each genre containing five different games with a total of approximately 20 thousand reviews. First, a pie chart was used to obtain the percentage of negative and positive feedback from each genre. The results that can be seen are that the five games with the First-Person Shooter (FPS) genre received a very low player satisfaction rate with a figure of 46.9%. On the contrary, games with the Indie genre have the smallest number of negative comments from players compared to other games, which is 8%. While for the other three genres, the negative reviews obtained were no more than 20%. From this, it can be concluded that the FPS genre tends to receive a lot of criticism from players, where this criticism may be related to aspects of graphics, storylines, or game performance such as smoothness and responsiveness of the game.

Next, a histogram graph was implemented to understand the distribution of review lengths for each genre. All five genres have relatively similar review lengths; most players provide concise feedback, but some players write very detailed reviews, possibly containing passionate opinions or in-depth criticism. The largest review length ranges from 18 to 55 characters across the five genres. The highest average review length is for FPS (495.14 characters) and the lowest for Indie (311.23 characters). Some conclusions that can be drawn from review trends over time are that FPS and Action (mean 399.24 characters) genres tend to have longer reviews, which may contain more detailed and nuanced feedback that is useful for further analysis. Since most of the reviews given are short, sentiment analysis will focus more on shorter texts.

Some offensive or inappropriate reviews that have been censored by the Steam Community have been found across various genres. After the pie charts were created, the information that can be derived is that the First-Person Shooter (FPS) genre has a censored review percentage of 10.6%, indicating the highest total usage of offensive or inappropriate language compared to other genres. This is followed by the Action genre, which has a percentage of 6.1%, making it the second highest for reviews containing censored text. On the other hand, the Strategy genre has the lowest percentage, at only 2.8%, reflecting relatively controlled language in the reviews for this genre.

Some players have requested refunds for the game. From the pie charts created, it can be seen that the percentage of players who requested refunds across the five genres does not exceed 10%, indicating that these figures are relatively low. The Action genre has 0.7% of players who have requested a refund, making it the highest genre in terms of players dissatisfied with the game they played and paid for. On the other hand, the Strategy genre has a figure of 0.2%, which is the smallest in this distribution.

Based on the graphical analysis of sentiment, censored text, and refund requests, it can be seen that the First-Person Shooter (FPS) genre received the most negative feedback from players. FPS recorded the highest percentage, which was 46.9% “not recommended,” the highest number of censored texts at 10.6%, and the second highest number of 0.5% for refund requests. In contrast, the Strategy genre received the most positive feedback among the five genres. This can be seen from the sentiment tone score which was the second smallest, which was only 10.0%, the smallest percentage of 2.8% for censored texts, and the smallest percentage of 0.2% for refund requests. From these results, it can be concluded that there is a possible relationship between dissatisfaction, inappropriate language, and product refunded.

4.3 Feature Engineering

Feature engineering in this chapter is the result of the data preparation phase, where the resulting features are considered useful for preparing the dataset for the machine learning phase.

1. Sentiment Tone Score

   In this project, VADER will be used to extract sentiment scores from the "Review Content" column, classifying reviews into positive, negative, or neutral tones. This feature serves as the foundation for sentiment classification, enabling deeper comparisons across different genres.

2. Review Length

   The Review Length feature provides information about the length or number of characters in each review submitted by players. This feature helps offer insights into whether longer reviews tend to be more positive or negative compared to shorter reviews.

3. Grouping of Playing Hours

   The Grouping of Playing Hours feature is categorized into three levels: low, medium, and high. This categorization simplifies the comparison of player satisfaction levels based on their playtime. It helps determine whether players with higher playtime are more likely to leave positive or negative reviews compared to those with lower playtime.

4. Time Pattern

The “Date Posted” column is transformed into a “Month-Year” format to facilitate the analysis of review trends over time. This feature helps identify periods of high review activity, which are likely linked to game updates, promotions, Christmas events, or other significant occasions.

5. Product Refunded

The Product Refunded feature is based on the players who requested a refund after playing the game. This feature is useful for analyzing players' dissatisfaction and finding out the root cause, such as bugs or the game not meeting the player's expectations. In addition, this feature can be combined with the Hours Played feature to see how much time players spent before deciding to request a refund.

6. Censored Text

Reviews containing offensive or inappropriate language, censored by Steam with the ♥♥♥ symbol, have been identified and analysed. This feature aims to capture the emotional nuances of such reviews and investigate whether reviews with censored language tend to be positive, neutral, or negative.

4.4 Machine Learning (Initial Results)

In this section, the initial approach used to apply machine learning techniques to analyze sentiment on Steam reviews is introduced. VADER (Valence Aware Dictionary and sEntiment Reasoner) is used to assign sentiment scores to reviews, which will later be part of the feature extraction process from the review text.

4.4.1 Feature Extraction

VADER predicts sentiment scores from the actual textual content of a particular review. These scores are divided into three numerical score components: positive, neutral, and negative. Additionally, a compound score that contains summarize the overall sentiment is also included.

	ID	Category	Review Content	Thumb Text	Sentiment	Compound	Positive	Neutral	Negative
0	814380	Action	one challenging rewarding game ever played.gre...	Recommended	Positive	0.8979	0.516	0.397	0.087
1	814380	Action	died twice	Recommended	Negative	-0.5574	0.000	0.217	0.783
2	814380	Action	best soul game . hesitation defeat .	Recommended	Neutral	0.0258	0.316	0.301	0.383
3	814380	Action	let known : wolf could kick malenia 's as .	Recommended	Neutral	0.0000	0.000	1.000	0.000
4	814380	Action	previously put 60 hour game ps4 , ready invest...	Recommended	Positive	0.9655	0.208	0.707	0.085

Figure 4.10: Example of Sentiment Scores Assigned by VADER

A portion of the datasets is processed using VADER to demonstrate how sentiment analysis transforms text into numerical sentiment scores. The sentiment scores for each review are stored as a new feature in the dataset. Figure 4.10 illustrates an example of how sentiment scores appear after processing.

4.4.2 Baseline Model

To illustrate how VADER output is used in sentiment analysis, the compound score is divided into three classes: positive, negative, and neutral, based on specific thresholds:

Compound score > 0.05: Positive

Compound score < -0.05: Negative

Compound score between -0.05 and 0.05: Neutral

From this, reviews will not only focus on whether they are positive or negative. This is because not all reviews given by players are entirely negative. There are some reviews where, even though players give a “not recommended” rating, they still write about both the good and bad things they experienced while playing the game. This makes the review neutral.

4.5 Conclusion

This chapter has discussed the initial insights gained from Exploratory Data Analysis (EDA) through graphical visualization. In addition, the features used in preparation for machine learning have also been described. Finally, the initial steps of implementing machine learning techniques have been presented, including initial results from the model used (VADER).