ANALYZING PLAYER FEEDBACK IN STEAM REVIEW ACROSS GAME GENRES

SAFIRA NURUL IZZA

UNIVERSITI TEKNOLOGI MALAYSIA

UNIVERSITI TEKNOLOGI MALAYSIA
DECLARATION OF Choose an item.

Author's full name	:	Safira Nurul Izza		
Student's Matric No.	:	MCS241009	Academic Session	:
Date of Birth	:		UTM Email	:
Project Report Title	:	ANALYZING PLAYER FEEDBACK IN STEAM REVIEW		

I declare that this thesis is classified as:

☒ **OPEN ACCESS** — I agree that my report to be published as a hard copy or made available through online open access.

☐ **RESTRICTED** — Contains restricted information as specified by the organization/institution where research was done. *(The library will block access for up to three (3) years)*

☐ **CONFIDENTIAL** — Contains confidential information as specified in the Official Secret Act 1972)

(If none of the options are selected, the first option will be chosen by default)

I acknowledged the intellectual property in the Choose an item. belongs to Universiti Teknologi Malaysia, and I agree to allow this to be placed in the library under the following terms :

1. This is the property of Universiti Teknologi Malaysia
2. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of only.
3. The Library of Universiti Teknologi Malaysia is allowed to make copies of this Choose an item. for academic exchange.

Signature of Student:

Signature :

Full Name
Date :

Approved by Supervisor(s)

Signature of Supervisor I:	Signature of Supervisor II
Full Name of Supervisor I NOOR HAZARINA HASHIM	Full Name of Supervisor II MOHD ZULI JAAFAR
Date :	Date :

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

This letter should be written by a supervisor and addressed to Perpustakaan UTM. A copy of this letter should be attached to the thesis.

Date:
Librarian
Jabatan Perpustakaan UTM,
Universiti Teknologi Malaysia,
Johor Bahru, Johor

Sir,

CLASSIFICATION OF THESIS AS RESTRICTED/CONFIDENTIAL
TITLE: Click or tap here to enter text.
AUTHOR'S FULL NAME: Click or tap here to enter text.

Please be informed that the above-mentioned thesis titled _____ should be classified as RESTRICTED/CONFIDENTIAL for a period of three (3) years from the date of this letter. The reasons for this classification are
(i)
(ii)
(iii)

Thank you.

Yours sincerely,

SIGNATURE:
NAME:
ADDRESS OF SUPERVISOR:

"Choose an item. hereby declare that Choose an item. have read this Choose an item. and in Choose an item.
opinion this Choose an item. is sufficient in term of scope and quality for the award of the degree of Choose an item."

Signature	:	_____
Name of Supervisor I	:	NA
Date	:	15 JANUARY 2025

Declaration of Cooperation

This is to confirm that this research has been conducted through a collaboration Click or tap here to enter text. and Click or tap here to enter text.

Certified by:

Signature :
Name :
Position :
Official Stamp
Date

* This section is to be filled up for theses with industrial collaboration

Pengesahan Peperiksaan

Tesis ini telah diperiksa dan diakui oleh:

Nama dan Alamat Pemeriksa Luar **:**

Nama dan Alamat Pemeriksa Dalam **:**

Nama Penyelia Lain (jika ada) **:**

Disahkan oleh Timbalan Pendaftar di Fakulti:

Tandatangan :

Nama :

Tarikh :

ANALYZING PLAYER FEEDBACK IN STEAM REVIEW ACROSS GAME GENRES

SAFIRA NURUL IZZA

A Project Report submitted in partial fulfilment of the
requirements for the award of the degree of
Master in Data Science

Choose an item.
Faculty of Computing
Universiti Teknologi Malaysia

JANUARY 2025

DECLARATION

I declare that this project report entitled *"Analyzing Player Feedback in Steam Review Across Game Genres"* is the result of my own research except as cited in the references. The project report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature	:	.....
Name	:	SAFIRA NURUL IZZA
Date	:	15 JANUARY 2025

ABSTRACT

This paper examines active player feedback on Steam, one of the most popular game distribution platforms, with the aim of identifying sentiment patterns across game genres. The binary review system used by Steam limits insight into player preferences, as reviews are only categorized as “Recommended” or “Not Recommended”. To address this limitation, this study applies sentiment analysis methods to 100,000 player reviews across five genres: Action, RPG, FPS, Strategy, and Indie. The data was obtained through automated web scraping, processed to ensure a consistent format, and analysed to uncover sentiment trends, review length, and other aspects. Initial results show significant differences in player expectations and satisfaction levels across genres. FPS games recorded the highest dissatisfaction rates, while Indie games had the highest satisfaction rates. Machine learning methods, such as VADER, as well as plans for the implementation of BERT, are proposed for sentiment classification to provide strategic insights useful to game developers. This study emphasizes the importance of integrating player feedback into the game design and development process to improve user experience and satisfaction. Further efforts include in-depth exploration of gameplay elements as well as the development of more sophisticated machine learning models to produce more accurate sentiment analysis.

ABSTRAK

Makalah ini mengkaji maklum balas pemain aktif di Steam, salah satu platform pengedaran permainan paling popular, dengan tujuan mengenal pasti corak sentimen merentasi genre permainan. Sistem ulasan biner yang digunakan oleh Steam menghadkan pemahaman terhadap keutamaan pemain, kerana ulasan hanya dikategorikan sebagai "Disyorkan" atau "Tidak Disyorkan". Bagi menangani batasan ini, kajian ini menggunakan kaedah analisis sentimen terhadap 100,000 ulasan pemain merangkumi lima genre: Aksi, RPG, FPS, Strategi, dan Indie. Data diperoleh melalui pengikisan web automatik, diproses untuk memastikan format yang konsisten, dan dianalisis untuk mengenal pasti trend sentimen, panjang ulasan, serta aspek lain. Hasil awal menunjukkan perbezaan ketara dalam jangkaan dan tahap kepuasan pemain merentasi genre. Permainan FPS mencatat kadar ketidakpuasan tertinggi, manakala permainan genre Indie mencatat kadar kepuasan tertinggi. Kaedah pembelajaran mesin seperti VADER, serta rancangan untuk pelaksanaan BERT, dicadangkan untuk pengelasan sentimen bagi menyediakan pandangan strategik yang berguna kepada pembangun permainan. Kajian ini menekankan kepentingan mengintegrasikan maklum balas pemain ke dalam proses reka bentuk dan pembangunan permainan untuk meningkatkan pengalaman dan kepuasan pengguna. Usaha seterusnya termasuk penerokaan mendalam terhadap elemen permainan serta pembangunan model pembelajaran mesin yang lebih canggih untuk menghasilkan analisis sentimen yang lebih tepat.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

FPS	-	First-Person Shooter
RPG	-	Role Playing Game
Indie	-	Independent video games
Colab	-	Google Colaboratory (Cloud-based coding and ML platform)
EDA	-	Exploratory Data Analysis
NLP	-	Natural Language Processing
API	-	Application Programming Interface
BERT	-	Bidirectional Encoder Representations from Transformers

CHAPTER 1

INTRODUCTION

1.1 Introduction

Steam is a successful game distribution platform, known as the place with the largest online video game sales transactions. According to Lanier (2019), out of a total of one billion registered accounts, around 90 million are active users. Every week, more than a thousand new games are released on Steam, which influences many of the factors that customers consider when deciding to purchase a game (Urriza & Clariño, 2021). Figure 1.1 shows the Steam homepage, where player can access games by genre, category, system requirement, and many more. However, Steam's binary review system, namely "Recommended" and "Not Recommended," limits the detail of player feedback. Often, players convey mixed sentiments that are difficult for developers to understand, even though this is very important for improving the playing experience and player satisfaction.

Moreover, it is hard to dividing games into genres as expectations of a player depend on the genre. Some are made based on realistic graphics, some can be based on the storyline, and some of them can be based on accurate game mechanics only. What makes it even more challenging is that the Steam review system is less detailed, and developers cannot analyse exactly which specific aspects that players like or dislike.

To overcome this research gap, this study applies a sentiment analysis approach and strains to derive further understanding from players' feedbacks. This way, game designing and development can be done with consideration of particular needs and preferences in each category. Therefore, this study proposes a computational-based method to identify trends and patterns that may be difficult to capture with conventional analysis.

Figure 1.1 Steam Website Home Page

1.2 Problem Background

Steam reviews are the primary way for game developers to understand what their customers think of their games. However, the binary technique used by Steam does not capture the complexities of how players feel. For example, a favourable review could simply say "It's okay, nothing special," and a negative review could say "Great concept, but too many bugs.". This lack of detail makes it difficult for creators to fully evaluate how their game is regarded, especially across genres. For instance, graphics, storyline, or control schemes, may be crucial features in one genre but insignificant in others. Lack of information about these genre preferences means that developers may miss out on ideas of how their applications could be better. Because players' expectations vary by genre, it's critical to go beyond review counts

However, the huge amount of review data generated annually makes it undeniably impractical to analyse all the feedback manually. For this reason, the utilization of automatic tools is mandatory in aiding with the extraction and processing of meaningful data from player reviews and feedback as a scalable approach to improving on subsequent games.

1.3 Problem Statement

In particular, the binary Steam review system is very limited, it allows offering only very general information about the player's attitude toward the game, it is either liked or disliked. Contemporary research does not pay much attention to the variation in the arrangements of the reviews by genre, which poses challenges to developers due to the lack of knowledge of the players' preferences across genres. Therefore, the intention of this study is to fill this gap by examining the sentiment trends of the players' reviews as a way of providing more information to game developers. More so, it aims at describing how various game genres are likely to be perceived, as well as establishing key features that would define sentiment, as well as any patterns that are credible in supporting development concerns.

1.4 Research Question

1. How can sentiment analysis of Steam reviews provide a deeper understanding of player feedback?

2. How does player feedback differ across various game genres based on reviews?

3. What specific elements of gameplay (e.g., story, graphics) do players tend to highlight in positive and negative reviews across game genres?

1.5 Aim and Objectives

The aim of this study is to examine the relationship between player feedback and game genres on Steam, uncover significant trends in review sentiments, and offer actionable recommendations for developers to enhance player satisfaction.

The objectives of these research are:

(a) To analyse the proportion of positive and negative reviews within each genre.

(b) To identify trends in player feedback across genres and provide actionable insights.

(c) To explore the relationship between specific game elements and player sentiments.

1.6 Scope of Study

This study employs sentiment analysis to evaluate the emotional tone and content of Steam reviews across five selected genres: Action, RPG (Role-Playing Game), FPS (First-Person Shooters), Strategy (Real-Time/Turn-Based), and Indie. Each genre is defined as follows:

1. **Action:** Action games emphasize fast-paced gameplay that requires physical challenges such as hand-eye coordination and quick reflexes. They often feature combat, platforming, or puzzle-solving elements and can overlap with other genres like action-adventure games (G2A.COM Editorial Team, 2024).

2. **RPG (Role-Playing Game):** Role-playing games allow players to assume the roles of characters in a fictional setting, often focusing on story-driven progression, character customization, and decision-making that impacts the game's world and narrative (Wieland, 2024).

3. **FPS (First-Person Shooters):** FPS games involve weapon-based combat viewed through the eyes of the player character, focusing on graphical fidelity, immersive environments, and gameplay mechanics that simulate combat scenarios (Sandmann, 2024).

4. **Strategy:** Strategy games require players to think critically and plan ahead to achieve objectives, often involving resource management, tactical combat, and long-term planning. These games can be categorized into turn-based or real-time strategies (Wakeham, 2024).

5. **Indie:** Games developed by individuals or smaller teams without the financial support of a large publisher, often characterized by innovative gameplay and unique artistic styles (as discussed in "The Rise of Indie Games," Patel, 2022).

Data will be sourced from five games within each genre, focusing on reviews written in English. The total dataset consists of 100,000 reviews to ensure robust analysis and meaningful conclusions. The study will utilize a combination of exploratory data analysis (EDA) and machine learning techniques to extract meaningful patterns and trends from the dataset. Advanced sentiment analysis models, such as VADER and BERT, will be employed to achieve high accuracy in detecting player sentiments.

1.7 Significance of Research

The study will try to fill the literature on differences in game genres and types of feedback received from players. Other than binary review systems, this study gives deeper insights into player preferences by genre. This approach enables developers to aim directly at what players mostly need, according to their feedback. For instance, since FPS players are more concerned with graphic quality and resolution than the storyline, by investing more into enhancing those factors, they create a far better experience for players.

This study also develops a novel method for the analysis of a large corpus of player reviews and game-related textual data that can easily be applied in other contexts or settings. The findings of this study enrich not only basic data but also provide a better theoretical understanding of the factors affecting player sentiment. Therefore, this research serves as an important reference for the developer and sentiment analysts of the game industry to make the development data-driven.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter aims to review the current state of knowledge regarding Steam review systems, sentiment analysis, data collection, and the platforms that will be used in the analysis process. This literature review also evaluates previous research methods that are relevant to the topic, so that it can help design the right method for this research as well as identify gaps in the literature that have not been widely explored.

2.2 Overview of Steam Review System

The review system of the Steam is categorized into two, Recommended or Not Recommended. From this simple approach, Steam exhibits clearly whether a particular game is relevant or not. However, this method reduces potentially much more varied opinions about individual players, for example, more intermediate options of recommendation, such as 'rather recommend,' or application of finer gradations of recommendation, for instance, by using a scale from 1 to 10. Guzsvinecz (2023) has also sought to look at how these systems impact the form of reviews and came up with dramatic differences between the two categories. It was also established that those tagged as "Not Recommended" contain negativity within the text and take longer elaborated texts to express it. On the contrary, positive reviews usually have fewer remarks, mostly which aspect that players enjoyed the most in a game. This is due to the limited choice of opinions in the Steam review system, the players give easily extreme opinions about it being either very satisfying, or very dissatisfying, with no room for those in the middle category.

2.3 Sentiment Analysis in User Review

Sentiment analysis is a suitable method for analysing and classifying emotions, understanding and processing textual data to see the sentiment contained in an opinion (Sari & Wibowo, 2019). Regarding the purpose of this method, it helps businesses and researchers understand user input, learn what people are saying regarding various issues, and overall improve the quality of decision making. When applied to the context of user reviews, sentiment analysis is useful for recognizing user satisfaction patterns and trends, such as their satisfaction and dissatisfaction with certain elements in a product or system. It refers to a specific branch within Natural Language Processing (NLP) and focuses on developing practical systems that can be used to extract opinions from text (Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Medan Area, 2022). Through this approach, previously unstructured information can be processed into more organized data.

For example, sentiment analysis can be applied to study player satisfaction on Steam by applying a data collection method to collect player review data. These reviews can later be classified into positive, neutral or negative sentiments, using several approaches which will be discussed in this chapter by reviewing several studies or papers that apply sentiment analysis.

2.4 Game Genres and Player Expectation

Players' expectations and feedback on games depend largely on their genre. RPG players, for example, prioritize in-depth storylines and character development, while FPS players tend to focus on graphic quality and gameplay mechanics. Based on a report from Statista (Clement, 2024), FPS will become the most played game genre worldwide in the second quarter of 2024, especially among players aged 16 to 54 years.

FPS games such as Call of Duty and Counter-Strike often offer a first-person perspective-based gaming experience, where graphics and controls are very important elements.

Meanwhile, the action-adventure genre is ranked second as the favourite genre in most age groups. This genre combines elements of real-time interaction and puzzle solving with interactive narrative, as found in popular games such as The Legend of Zelda and Grand Theft Auto. This difference shows how RPGs are often criticized on the narrative aspect, while FPS receive more attention on the visual and gameplay aspects. This emphasizes the importance of a specific analytical approach to understanding the unique expectations in each genre.

Although game genres have a large influence on player responses, research examining sentiment trends across genres is still relatively minimal. According to Guzsvinecz and Szabó (2023), many studies only focus on general feedback, without paying attention to players' unique preferences in each genre. These shortcomings highlight the importance of more targeted sentiment analysis to understand and meet the specific expectations of each genre.

2.5 Sentiment Analysis Method

There are different types of sentiment analysis that can be used to identify user responses. Several sentiment analysis methods will be discussed based on studies and research that have been carried out.

2.5.1 Lexicon-Based Methods

A predefined sentiment dictionary is used by lexical-based methods to determine the meaning tendencies of words in a text. The dictionary results are then combined to calculate the overall sentiment of a sentence. This method is simple to implement and works well with short texts but lacks the ability to understand context or sarcasm (Railean, 2024).

2.5.2 Machine Learning-Based Methods

Machine learning methods, such as SVM or Naive Bayes, learn patterns from categorized datasets to classify sentiments. These methods are robust for classification but require significant preprocessing and labelled data (Aloufi & El Saddik, 2018; Pai & Liu, 2018).

2.5.3 Deep Learning-Based Methods

Deep learning models, such as LSTMs, CNNs, or Transformers, use neural networks to analyze text, capturing context, word dependencies, and nuances. This method is effective for processing data on a large scale, but involves a complex calculation process that requires large computing resources. (Alaparthi & Mishra, 2020; Kokab et al., 2022).

2.5.4 Hybrid Methods

Hybrid methods combine lexicon-based methods, machine learning or deep learning and utilize the advantages of each method to achieve a specific goal. Although this method is flexible and powerful, it still requires careful adjustments so that all methods can work well (Ahmed et al., 2022; Novel Approach, 2019).

2.5.5 Aspect-Based Sentiment Analysis (ABSA)

Aspect-Based Sentiment Analysis or ABSA is an attempt to determine sentiment towards aspects of a product or system, for example game play or graphics in game reviews. This approach offers more information, but labelled data relating to each aspect must be used (Jiang et al., 2023; Wang et al., 2019).

Table 2.1 Comparison of Sentiment Analysis Methods, Applications, and References

Method	System / Application	References
Lexicon-Based Methods	Used for sentiment analysis in general text.	Railean, 2024
Machine Learning-Based Methods	SVM applied to domain-specific data like football tweets and vehicle sales predictions.	Aloufi & El Saddik, 2018; Pai & Liu, 2018
Deep Learning-Based Methods	Deep learning models like Transformers and BERT for social media sentiment analysis.	Alaparthi & Mishra, 2020; Kokab et al., 2022
Hybrid Methods	Hybrid approaches integrating lexicon-based and machine learning for multilingual corpora.	Ahmed et al., 2022; Novel Approach, 2019
Aspect-Based Sentiment Analysis (ABSA)	Aspect-based sentiment analysis for product reviews, targeting specific aspects like features or performance.	Jiang et al., 2023; Wang et al., 2019

2.6 Data Collection Method

Motivated by the work of Ahmed et al. (2022) on multilingual sentiment analysis, it has been observed that the quality and variety of data is critical to make sentiment analysis effective. The ways and means of data collection for sentiment analysis about a particular product, brand, company or any entity of interest have undergone a lot of changes with the enhancement of web technologies and availability of data (Chen & Zhang, 2020). Recent literature outlines three commonly used approaches for data collection: web scraping, APIs and data set repository (Jiang et al., 2023; Pai & Liu, 2020).

2.6.1 Web Scraping

Web scraping as the name suggests is the process of collecting data existing on the websites often with the help of computer scripts. This method is very helpful when it comes to capturing high volumes of free text from the web blogs, forums and ecommerce sites and the like. For instance, Chen and Zhang (2020) conducted web scraping of Amazon to gather user reviews before using them to assess sentiment trends in product feedback. Flexibility is the key advantage of web scraping, but practice must adhere to the provided ethical rules and regulations of the website to prevent legal trouble or invading privacy of people (Jiang et al., 2023).

2.6.2 APIs

Application Programming Interfaces (APIs) provide a systematic means of gathering information from different networks such social media sites, news sources, and review sites. APIs are preferred mainly because they are easy to integrate and because they are highly dependable in providing structured data. For instance, Aloufi and El Saddik (2019) collected football related tweets which allowed them to perform domain specific sentiment analysis by using the Twitter API. However, some restrictions include API level, which may reduce the extent of utilization in large-scale projects besides costs of access (Pai & Liu, 2020).

2.6.3 Dataset Repository

A dataset repository is data that has been gathered in advance in most cases and accumulated to serve a research study. Consequently, this repository is very helpful in a way that it provides labelled datasets which are very useful when developing and testing sentiment analysis models. While IMDB annotations and sentiment analysis on the 140-character Twitter comments are typical for general sentiment analysis, there are numerous repositories by specific domains. In their survey, Ahmed et al. (2022) also stress that multilingual sentiment analysis systems have to rely on dataset repositories. However, data in a repository might not always be up to date indicating the need for other means of data acquisition.

Table 2.2 Comparison of Data Collection Methods

Data Collection Method	Purpose	Use Case
Web Scraping	Collects unstructured data from websites for sentiment analysis.	Extracting reviews from e-commerce platforms.
APIs	Structured access to data from platforms like social media.	Collecting tweets via Twitter API.
Dataset Repositories	Provides pre-labeled data for training and testing models.	IMDB movie reviews for general sentiment.

2.7 Tools and Platforms

Google Colaboratory and Jupyter Notebook are two widely used tools for data analysis and machine learning. Their features and limitations are summarized in the table 2.3.

Table 2.3 Comparison of Google Colaboratory and Jupyter Notebook

Feature	Google Colaboratory	Jupyter Notebook
Platform	Cloud-based, accessed via a web browser.	Local or server-based installation.
Setup	Pre-configured with libraries like TensorFlow, Keras, PyTorch, and OpenCV.	Requires manual installation of libraries.
Hardware Resources	Provides free access to GPUs and TPUs.	Relies on user hardware. Can integrate with GPUs if available.
Performance	Suitable for GPU-centric tasks, less efficient for CPU-based operations due to limited cores.	Performance depends on local machine specs; scales with better hardware.

Feature	Google Colaboratory	Jupyter Notebook
Runtime Limitations	Session lasts up to 12 hours, with GPU usage potentially restricted after extended use.	Unlimited runtime, contingent on local hardware capabilities.
Cost	Free of charge with Google Drive integration.	Cost depends on user hardware and maintenance.
Collaboration	Integrated sharing via Google Drive; supports real-time collaboration.	Limited sharing options; third-party tools required for collaboration.
Data Handling	Seamless integration with Google Drive.	File handling depends on the local environment or additional setups.
Offline Access	Not available, requires internet connection.	Fully functional offline.
Use Case	Ideal for quick prototyping, education, and projects requiring GPU access without dedicated hardware.	Preferred for projects needing full control over the environment or long-term computations.

Both tools have their strengths and serve different user needs. Google Colaboratory is great for users with limited resources or those needing quick setup and GPU support. Jupyter Notebook is better for advanced users who want more control and offline access. Using both tools together can create an efficient workflow. This summary is based on Carneiro et al. (2024).

2.8 Challenges in Sentiment Analysis

Although sentiment analysis has now emerged as one of the important tools for measuring user opinions and feedback in various applications, there are still several issues that hinder the level of accuracy and flexibility of the process. This is due to the nature of natural language, limitations of current approaches, and issues related to sample collection and modelling. This section presents basic knowledge about some of the most frequently occurring sentiment analysis problems.

2.8.1 Handling Sarcasm

Sarcasm is a form of sentiment and is the most complicated factor in sentiment analysis because it requires comparing the hidden meaning with the words written explicitly. In most previous sentiment analysis models, sarcasm is an aspect that often goes unnoticed, resulting in misinterpretation of words. However, much recent research has led to the development of complex models to solve this problem. For example, Vitman, Khmelevskiĭ, and Semenova (2022) proposed a contextual framework, which leverages context, affective information, and sentiment to improve sarcasm detection for SM messages. In addition, Kaseb and Farouk (2023) proposed a more accurate system in terms of sentiment, sarcasm, and dialect for Arabic tweets because it considers the elements of sarcasm in tweets. These studies have revealed that the context and emotional information combined play a role as key factors to help eliminate the difficulties associated with sarcasm in sentiment analysis models.

2.8.2 Ambiguity in Language

One of the major obstacles in sentiment analysis comes from the use of words with multiple meanings, in other words, polysemy. This can be confusing to sentiment analysis systems that does not factor in context into the equation. For instance, the word “cool” within the context of a text may be viewed to mean something positive, but on the other hand, it can also refer to temperature. Another benefit of the model is handling such feature complexity, and other similar systems include the BERT transformer. Yet, the most favorable performance is reached when working with large datasets that comprehensively labelled (Alaparthi & Mishra, 2020).

2.8.3 Multilingual and Cultural Differences

When it comes to multi-languages and diverse cultures, sentiment analysis becomes a little more difficult. Often, words and expressions have different connotations in different cultures making it difficult to establish models for different cultures. For example, Ahmed et al., 2022 outlined that a systematic combination of approaches is needed in the context of sentiment analysis in different languages.

However, the available labelled data is insufficient for many languages making it one of the biggest challenges in this field.

2.9 Conclusion

This chapter includes a review of the literature, different research on sentiment analysis and analysis applied in various fields. This chapter highlights various sentiment analysis methods, and data collection methods followed by examples from different studies. Apart from that, this literature review also presents the difficulties experienced in doing sentiment analysis, for example, how to differentiate between sarcasm, handle figures of speech with varying meanings, or overcome the barriers of different languages and cultures. When applied to these challenges, each analysis method has different benefits and approaches.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter starts by explaining the research framework to conduct sentiment analysis on Steam player reviews across genres. Furthermore, the dataset collection method is also explained in this chapter. After the dataset collection process was completed, the data went through a preprocessing stage, where data cleaning and transformation were performed to produce relevant data for analysis. Finally, this chapter provides an explanation of feature engineering to prepare the features needed in machine learning based on the findings from the data processing process that has been carried out.

3.2 Research Framework

There are five phases in the research framework to analyse player review sentiment on Steam games across genres, as shown in figure 3.1. The first phase is the identification of research gaps, where various studies and papers relevant to this project have been reviewed and studied. Through this phase, insights are obtained to apply the best method or model that can be implemented into this project. In addition, the identification of research gaps also aims to find aspects that have not been widely discussed in previous studies, such as sentiment analysis on reviews containing censored text or the label "product refunded". Phase 1: Research Gap Identification was conducted and discussed in Chapter 2 where relevant literature was reviewed for this study.

The next phase is data collection, where review data is taken and gathered directly from Steam Community Reviews. This ensures that the dataset used is original

data that comes directly from players. The third phase is data preprocessing, where raw data is cleaned and transformed into a consistent and structured format suitable for analysis. The fourth phase is Exploratory Data Analysis (EDA), aimed at understanding the structure, trends, and patterns hidden within the data. The next phase is Feature Engineering, which involves creating or modifying features in the dataset to improve the performance of machine learning models. The final phase is machine learning preparation. This phase consists of two processes: feature extraction and model setup. The goal of feature extraction is to transform the cleaned and pre-processed text into a numerical format. After that, the feature extraction process continues to separate the data and select the machine learning model

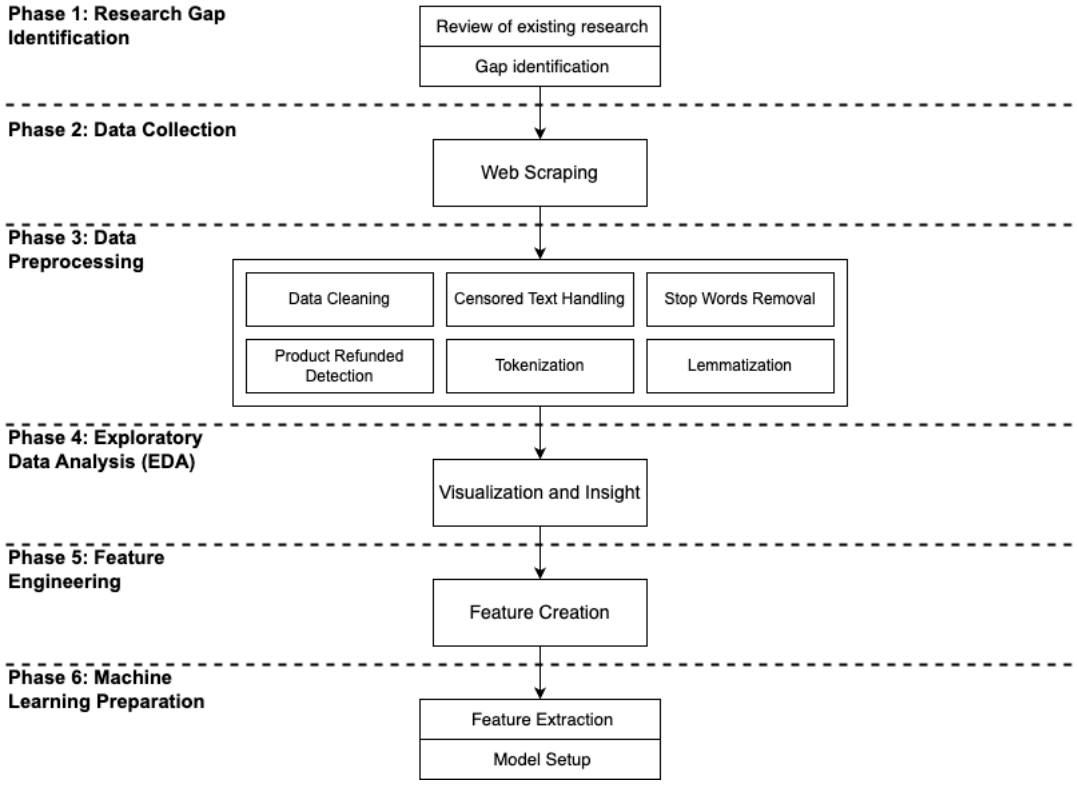

Figure 3.1 Research Framework

3.3 Data Collection

Continuing the research framework, the next step is Phase 2: Data Collection. Determining the data source is a crucial step in the sentiment analysis process. In this project, it is essential to ensure that the review data used is original and obtained directly from a trusted source, which in this case is the players themselves, without any modifications or manipulation. Once the data source is determined, the next step is to proceed with data collection. During the data collection phase, the relevance of the data to the project objectives was also considered to ensure that the analysis provided results aligned with the project's goals.

3.3.1 Data Source

The data source for the analysis process was taken from the Steam Community that is available on the official Steam website, accessible via the link https://steamcommunity.com/. The Steam Community is a discussion and interaction platform for players, which provides user reviews of games they have played. Each review on the Steam Community includes information such as the username, the content of the review, an indication of whether the player recommends the game (Recommended/Not Recommended), the amount of time spent playing, and the date the review was published.

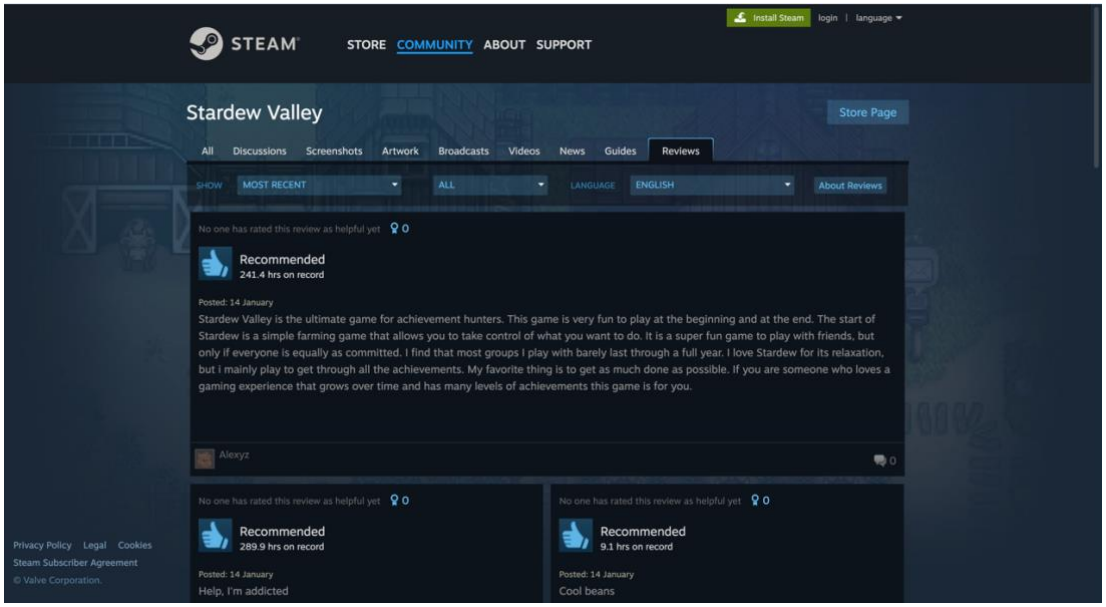

Figure 3.2 Steam Community Review Page

To ensure sufficient review data collection for analysis, the games selected for this study are those with large and active communities on the Steam platform. These games were chosen from the very active communities of players with high engagement for rich and diverse review data to support in-depth sentiment analysis and cover a wide variety of genres to be representative of the trend of sentiment within different gaming experiences. Table 3.1 lists the selected games, categorized by genre, along with their Steam IDs.

Table 3.1 List of the Selected Games

Title	Game Name	Game ID
Action	Red Dead Redemption 2	1174180
	Sekiro™: Shadows Die Twice - GOTY Edition	814380
	Lies of P	1627720
	Mortal Kombat 1	1971870
	God of War	1593500
First-Person Shooter	Call of Duty®	1938090
	Borderlands 3	397540
	Metro Exodus	412020

Title	Game Name	Game ID
	PUBG: BATTLEGROUNDS	578080
	S.T.A.L.K.E.R. 2: Heart of Chornobyl	1643320
Role Playing Game	DAVE THE DIVER	1868140
	Fear & Hunger	1002300
	Fallout 4	377160
	DREDGE	1562430
	Persona 5 Royal	1687950
Strategy	Balatro	2379780
	Dwarf Fortress	975370
	Hearts of Iron IV	394360
	Planet Zoo	703080
	Age of Empires II: Definitive Edition	813780
Indie	Undertale	391540
	Mouthwashing	2475490
	Unpacking	1135690
	MiSide	2527500
	Stardew Valley	413150

3.3.2 Selenium WebDriver

Since the Steam Community platform requires manual scrolling down to load additional reviews, Selenium WebDriver is used to automate this process. Selenium is a popular open-source tool for automating web browsers. With Selenium, we can interact with web elements, simulate user actions (such as clicking or scrolling), and automatically run tasks on web applications. One of the main advantages of Selenium is that it supports multiple browsers, so it can be used on different types of browsers without any problems (Kumari, 2024).

In Chapter 2, it explained how to choose a data collection method by reviewing the literature to find the best tools and platforms for efficient and reliable web scraping.

Google Colab was chosen because it is cloud-based, so it can run and process scripts faster than local platforms such as Jupyter Notebook.

However, when Google Colab is used for web scraping, this platform cannot connect directly to WebDriver installed on the local computer, even though all settings and connections have been confirmed to be connected correctly. To overcome this problem, Jupyter Notebook was finally used as an alternative. By running on a local computer, Jupyter Notebook can connect to WebDriver and run the web scraping process without compatibility issues. Although Google Colab is faster, Jupyter Notebook was chosen because of its compatibility with the required tools.

Several browsers that support WebDriver have been tested for the web scraping process, including Google Chrome and Microsoft Edge. Initially, ChromeDriver was selected and used for the web scraping process. However, after a version update on the Chrome browser, ChromeDriver was no longer compatible due to the version difference between the browser and WebDriver. As a result, the WebDriver had to be replaced with Microsoft Edge WebDriver.

```
from selenium.webdriver import Edge, EdgeOptions
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
```

Figure 3.3 Selenium WebDriver using Microsoft Edge WebDriver

3.3.3 Data Collection Method

The reviews were collected through web scraping using Microsoft Edge WebDriver using Jupyter Notebook as the platform. Data for 5 genres (Action, FPS, Indie, RPG, and Strategy) were collected, consisting of 20,000 reviews per genre, resulting in a total dataset of 100,000. The scraping process targeted certain attributes, such as review content, thumb text (recommended or not recommended), play hours, and date posted. To maintain consistency, review filters were applied to only include

reviews written in English. Additional filters were also applied, such as "Most Helpful" and "Show All Reviews." The "Most Helpful" filter displays reviews that other players find relevant or relatable. Figure 3.4 shows the filters that were set for the data collection process.

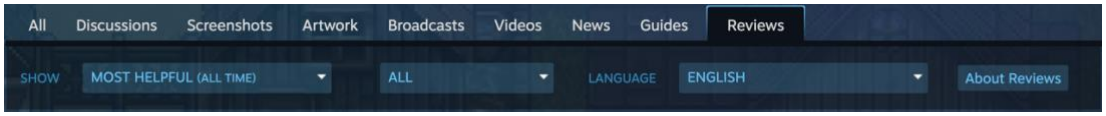

Figure 3.4 Steam Review Filters

Figure 3.5 shows examples of two reviews that received a lot of agreement and engagement from other players. One review, written by a user with the display name **S A L S A** ⧘, states: "*You play this game every night on your bed when you imagine fake scenarios like having your own house, money, and a wife.*" This review was voted helpful by 1,231 people, showing its impact in the community.

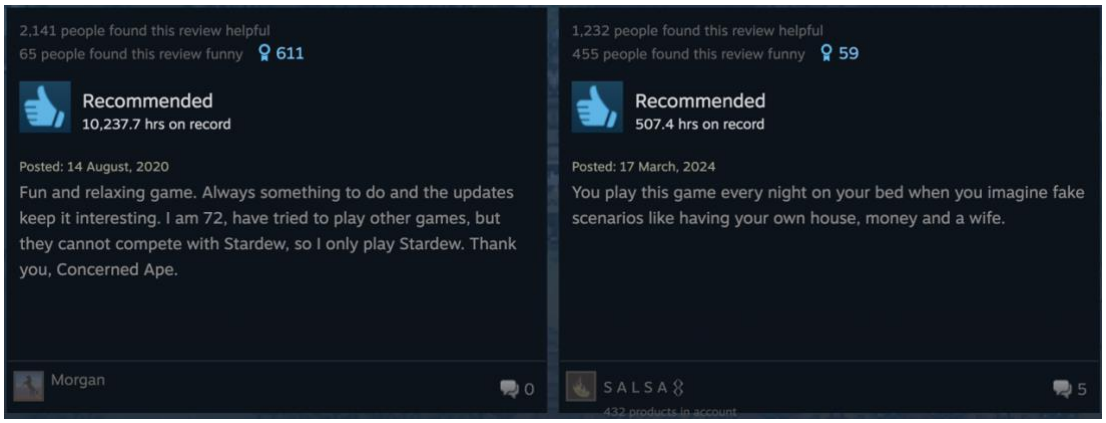

Figure 3.5 Examples of Highly Engaged Reviews

The "Show All Reviews" filter was used to display all types of reviews, regardless of whether the review is Recommended or Not Recommended. By activating this filter, the data collection process becomes more comprehensive because it includes reviews from both sides of the player's perspective, both those who provide positive recommendations and those who provide criticism or complaints.

Initially, the number of reviews required for each game in the dataset was set at 4,000. However, to avoid significant data reduction during the data cleaning process, the number of reviews retrieved was increased to 5,000 per game. Nevertheless, during the scraping process on some games, there was an obstacle where the number of reviews with the "Most Helpful" filter did not meet the target of 5,000 reviews. To overcome this problem, the scraping process was repeated once again on games that did not meet the target using the "Most Recent" filter. In this second scraping process, additional code was applied to ensure that the retrieved reviews were not duplicated with reviews from the "Most Helpful" filter. After the target number of reviews was reached, the two groups of reviews from different filters were merged into one dataset.

3.4 Data Preprocessing

Data preprocessing is a crucial step aimed at preparing raw data for further analysis. Raw data collected through web scraping often contains incomplete data (e.g., missing values or empty data), inconsistencies, and information that is irrelevant to the study's objectives. These issues can reduce the accuracy of the analysis. The data preprocessing process ensures that the dataset is clean and consistent before moving on to the analysis phase.

3.4.1 Data Cleaning

Data cleaning is done to ensure that the dataset is relevant, accurate, and ready for analysis. Figure 3.6 shows how the review cleaning flows from raw data to data that is ready to be processed in the next stage. This data cleaning process was carried out in Python, utilizing the pandas library on the Google Colab platform.

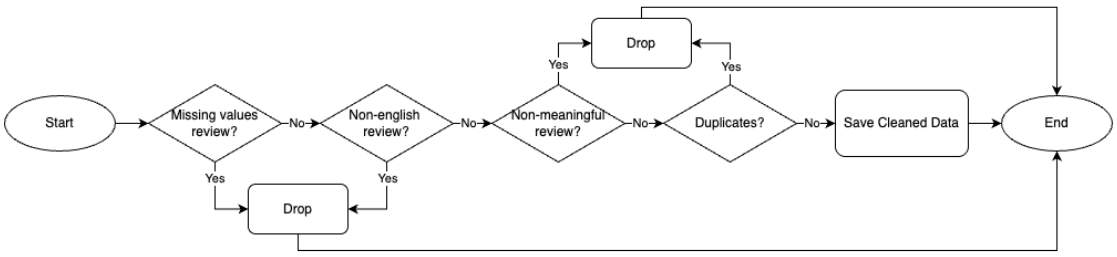

Figure 3.6 Data Cleaning Flow

The process begins by checking for missing values in the Review Content, Thumb Text, Review Length, Play Hours, and Date Posted columns. Missing values in any of these five columns can make the analysis process less optimal. After that, reviews that are not in English are also removed, because this analysis only focuses on English-language sentiment analysis. Although during data collection it was set to only collect English-language reviews, there were still some reviews written in other languages. Therefore, removing reviews that are not in English is an important step to ensure consistency and accuracy in sentiment analysis.

Not all players provide detailed and long feedback; there are some reviews that only contain one word, such as "Great," "Nice," "Bad," or "Foul." This type of review is less helpful in the analysis process because of the lack of meaningful feedback. Therefore, reviews that are less than six alphanumeric characters long are removed. Because the required dataset must have data variations, a check is carried out on duplicate review content. If the program detects a duplicate review, the review is deleted.

Several reviews were found to contain love emoticons, as seen in Figure 3.7. At first glance, one might think that players are using love emoticons because of the positive experience they got from the game. However, upon further examination by reading the overall context of the review, it can be ascertained that the emoticons are replacements for words that have been censored by Steam. It turns out that Steam replaces inappropriate words or swear words in player reviews with the ♥♥♥ emoticon.

Figure 3.7 Reviews With Censored Words

Not all reviews containing censorship are bad reviews. It could be that it is a way for players to convey their positive feedback with excitement or strong emotions towards their playing experience in the game. To maintain consistency in sentiment analysis, all censored words with emoticons were changed to "[censored]" for further analysis. Some reviews had the label "Product Refunded.", as illustrated in Figure 3.6. Reviews with this label were identified and retained. This is important for analysing player dissatisfaction patterns and understanding the reasons behind game refunds.

Figure 3.8 Review Labelled as "Product Refunded"

3.4.2 Data Transformation

After cleaning the data, certain attributes were standardized to ensure consistency and enable meaningful analysis. To ensure clarity and structured analysis, categories (genres) and game IDs were added to the dataset. This step aimed to group data based on games and genres as unique identifiers, so that data integrity could be maintained properly. Next, the `Play Hours` and `Date Posted` columns were standardized to maintain consistency across datasets. This process was useful because the data became simplified, making the analysis, visualization, and calculation processes easier.

	Review Content	Thumb Text	Review Length	Play Hours	Date Posted
0	the closest we're getting for a bloodborne gam...	Recommended	57	36.7 hrs on record	Posted: 29 October, 2023
1	the children yearn for bloodborne	Recommended	29	26.5 hrs on record	Posted: 3 February
2	Experience the horror of being french	Recommended	32	23.3 hrs on record	Posted: 15 October, 2023
3	all right then. keep your Bloodborne Sony.	Recommended	36	101.6 hrs on record	Posted: 21 December, 2023
4	They really went fine i'll make Bloodborne on ...	Recommended	50	41.5 hrs on record	Posted: 14 October, 2023

Figure 3.9 Raw Review Data Before Standardization

	ID	Category	Review Content	Thumb Text	Review Length	Play Hours	Month-Year
0	1174180	Action	the closest we're getting for a bloodborne gam...	Recommended	57	36.7	10-2023
1	1174180	Action	the children yearn for bloodborne	Recommended	29	26.5	02-2024
2	1174180	Action	Experience the horror of being french	Recommended	32	23.3	10-2023
3	1174180	Action	all right then. keep your Bloodborne Sony.	Recommended	36	101.6	12-2023
4	1174180	Action	They really went fine i'll make Bloodborne on ...	Recommended	50	41.5	10-2023

Figure 3.10 Review Data After Standardization

The next step was to convert all text to lowercase to ensure uniformity and avoid case-sensitivity issues. This process aims to simplify text processing for machine learning models. Commonly used words such as “the,” “a,” “an,” or “in” are removed using the Natural Language Toolkit (NLTK). This is done to ensure the analysis focuses only on important words, thereby improving the accuracy and efficiency of sentiment analysis in a given text.

Next, tokenization was done to break down a text into small units or part, that is similar to what we call tokens. This process aims to change unstructured text into

structured data so that it is easier to analyse. After tokenization, the lemmatization step was performed to reduce each word to its basic form (root word or lemma), while maintaining its meaning and context. Both of these processes are very important in sentiment analysis, because computers process data in numerical form, not raw text. After the dataset went through these processes, the transformed dataset was saved for further analysis.

3.4.3 Feature Engineering

Feature Engineering is the process of creating or changing features in datasets to make a machine learning model perform better. To get maximum results, the quality of the features used greatly determines the success of the machine learning model, so that it can find important patterns and relationships in the data so that the model can learn better. (GeeksforGeeks, 2023). In this project, several features are created to extract more meaningful information from the data that has been collected and has gone through various processes. These features can be seen below: through various processes. These features can be seen below:

1. Sentiment Tone Score

   The sentiment score was taken from the `Review Content` column using one of the sentiment analysis tools, where VADER will be used in this project . This score is used to determine whether the tone of the review is positive, negative, or neutral, which is the basis for sentiment classification.

2. Review Length

   A new feature called Review Length has been created to measure the length of each review based on the number of characters. This feature helps provide insight into the correlation between the number of characters in player feedback and sentiment.

3. Grouping of Playing Hours

   The hours spent by players in a game have been categorized into three labels, low, medium, and high. This labelling is useful for providing further understanding regarding the level of player engagement, and getting a comparison between players' satisfaction with the game and their playing time.

4. Time Pattern

   The `Date Posted` column is changed to `Month-Year` format for analysing review trends over time. This transformation makes it easier to create visual graphs and compare review patterns over a period of time.

5. Product Refunded

   A feature called Product Refunded has been created to identify game products that have been returned by players. This feature was created because during the data cleaning process, several games were found to have the label “Product refunded”. This makes the label a valuable feature to get the discovery of player dissatisfaction so that they want to return the product.

6. Censored Text

   A unique attribute was found in the dataset, where text containing inappropriate or offensive language in Steam reviews was censored using the ♥♥♥ emoticon. A feature called Censored Text was created to capture the intensity and emotional sentiment expressed in these reviews.

The six features mentioned above are important for understanding player behaviour, identifying sentiment trends, and being well prepared for the machine learning process

3.5 Conclusion

This chapter outlined the methodology employed to conduct sentiment analysis on Steam player reviews across genres. Starting with Phase 1, the research identified gaps in existing studies through a literature review, as discussed in Chapter 2. Next, the methodology moved to data collection and data preprocessing phases to ensure clean, structured and meaningful datasets. In addition, feature engineering was carried out to enhance the analytical potential of the datasets.

CHAPTER 4

INITIAL RESULTS

4.1 Introduction

This chapter provides an explanation of Exploratory Data Analysis (EDA) and the application of the dataset to eight EDA processes. Feature engineering is also discussed to create and modify features or variables in the input data space to improve machine learning outcomes. In addition, this chapter discusses feature extraction that converts textual review data into numeric form, so that reviews can be combined with other numeric features. Finally, a machine learning model is presented to provide initial recommendations.

4.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an important process for understanding data. EDA is used to find patterns, find unusual data, test ideas, and make assumptions using simple statistics and graphs (Patil, 2018). There are several steps in implementing EDA to ensure a good understanding of the data. The first is to understand the problem to be solved and the data available. Next is check the structure of imported data to find missing values and inconsistencies. Missing values need to be handled properly, either through imputation (filling in data) or deletion, to avoid biasing the analysis.

Next, data characteristics such as distribution, mean, and variance are explored to find patterns and outliers. Data transformations, such as scaling, encoding, or merging, are performed to make the data ready for analysis. Data visualization is then used to reveal relationships and trends through graphs and charts. Outlier management is also important to ensure the results of the analysis are credible. Finally, findings and

insights are clearly conveyed using visual aids and summaries that highlight key findings and potential further research (GeeksforGeeks, 2025). The flow of EDA processes can be seen in figure 4.1.

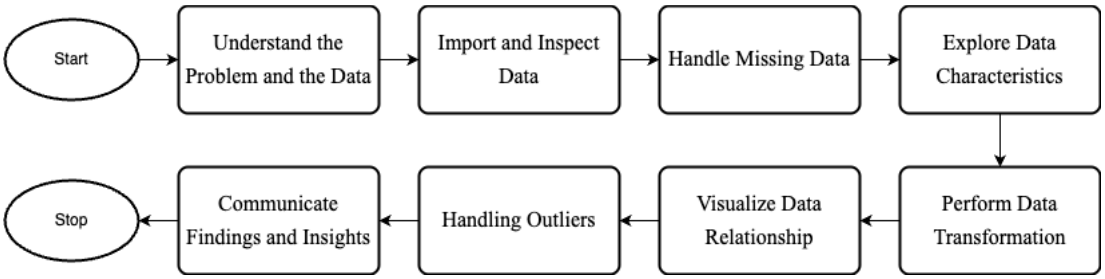

Figure 4.1 EDA Processes

4.3 Steps of Exploratory Data Analysis

The 8 steps of EDA will be discussed and implemented using the 100,000 player review data that has been collected.

4.3.1 Understand the Problem and the Data

Following the EDA steps, the first step is Understand the Problem and the Data. From the third phase of the Research Framework, which is data preprocessing that has been explained in Chapter 3, the data cleaning and data transformation processes have been carried out. Some of the steps implemented include removing missing values and duplicate values, changing the value format in the `Play Hours` column from object type to float or decimal, and changing the format of the `Date Posted` column to `Month-Year`. In addition, a new column was added to mark reviews containing censored words, a new column for reviews from players who refunded the game, changing the text in the `Review Content` column to lowercase, and removing stop words.

The final results of the data preprocessing phase can be seen in Figure 4.2, that shows 5 rows of data from one genre, namely the action genre. The `ID` column shows the game ID on Steam, `Category` shows the game genre, `Review Content` is the review of the experience given by players about the game, where this column has gone through the process of removing stop words and being changed to lowercase. The `Thumb Text` column shows whether the player likes or dislikes the game, `Review Length` is the length of the original review calculated based on the number of characters, `Play Hours` shows the duration of the player's playtime (in hours), and `Month-Year` is the month and year when the review was posted. In addition, there are additional columns, namely `Is_Refunded` to find out whether the review comes from a player who refunded the game, and `Contains_Censored` to find out whether the review contains censored words.

	ID	Category	Review Content	Thumb Text	Review Length	Play Hours	Month-Year	Is_Refunded	Contains_Censored
0	814380	Action	one challenging rewarding game ever played.gre...	Recommended	110	47.9	12-2024	False	False
1	814380	Action	died twice	Recommended	17	58.3	09-2023	False	False
2	814380	Action	best soul game . hesitation defeat .	Recommended	36	88.7	11-2024	False	False
3	814380	Action	let known : wolf could kick malenia 's as .	Recommended	39	46.9	07-2024	False	False
4	814380	Action	previously put 60 hour game ps4 , ready invest...	Recommended	1249	35.9	04-2024	False	False

Figure 4.2 Action Genre Datasets

4.3.2 Import and Inspect Data

In the second step of EDA, it is important to identify the data type of each variable because it will help in the next data manipulation and analysis steps. In Figure 4.3, the total columns in the dataset are 9. The `ID` and `Category` columns will not be used in the analysis process because both only function as game identities. The `Month-Year` column shows the time when the review was posted by the player, so the data type for this column should be changed to datetime format.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20000 entries, 0 to 19999
Data columns (total 9 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   ID                20000 non-null  int64
 1   Category          20000 non-null  object
 2   Review Content    19999 non-null  object
 3   Thumb Text        20000 non-null  object
 4   Review Length     20000 non-null  int64
 5   Play Hours        20000 non-null  float64
 6   Month-Year        20000 non-null  object
 7   Is_Refunded       20000 non-null  bool
 8   Contains_Censored 20000 non-null  bool
dtypes: bool(2), float64(1), int64(2), object(4)
```

Figure 4.3 Action Genre Datasets Information

Still in figure 4.3, it can be seen that some columns in the data used int64 and float64 types. Although these types ensure high precision, they can consume more memory than necessary. Therefore, down casting was performed on these columns to reduce memory usage and improve processing efficiency.

Figure 4.4 shows the code used to convert the `Month-Year` column data type to datetime format and to downgrade the `Play Hours` and `Review Length` columns data types to 32-bit. This code was written using Python and Pandas libraries. The results of this process are shown in Figure 4.5.

```python
import pandas as pd

# Convert Moth-Year column data type to datetime format
df['Month-Year'] = pd.to_datetime(df['Month-Year'], format='%m-%Y')

# Downcast the Play Hours and Review Length columns
df['Play Hours'] = df['Play Hours'].astype('float32')
df['Review Length'] = df['Review Length'].astype('int32')
```

Figure 4.4 Codes of Datetime Conversion and Data Type Down Casting

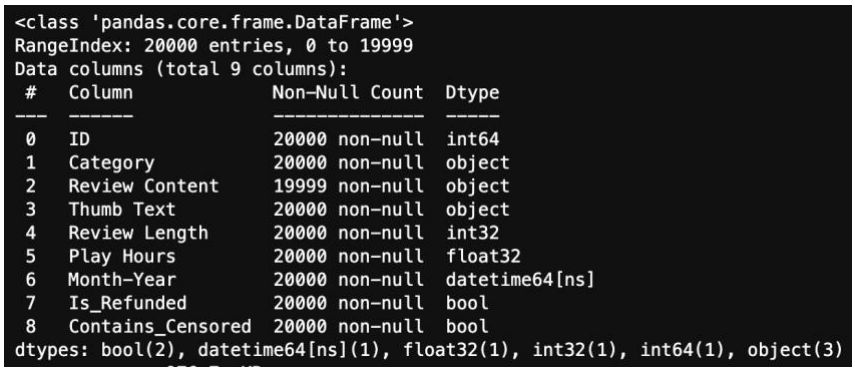

Figure 4.5 Action Genre Datasets Information After Conversion and Down Casting

4.3.3 Handle Missing Data

Although the dataset has gone through the data cleaning process in the data preprocessing phase, it turns out that null or missing values are still found in the `Review Content` column in all genres. This likely happened after the stop words removal process. To overcome the missing data in the `Review Content` column, the problematic data will be deleted. The reason of the data removal is because analysis can only be done if there are reviews that can be processed, so this column is very important. The code used to overcome this problem can be seen in Figure 4.6.

```
df = df[df['Review Content'].notna()]
```

Figure 4.6 Code to Drop Missing Value

4.3.4 Explore Data Characteristics

After dealing with missing data, the next step is to explore the characteristics of the dataset to understand its central tendency, distribution, and variability. Summary statistics are calculated for numeric variables, and visualizations are created to help identify patterns and anomalies in the data (GeeksforGeeks, 2025).

The numeric variables to be calculated are the hours played for each genre, using a box plot. As shown in figure 4.7, The box plot shows outliers in the `Play`

`Hours` column, especially in the Strategy and FPS genres. These outliers likely reflect highly engaged players or are simply unusual data. Additionally, the skewness values show that most genres have a distribution that is skewed to the right, indicating the presence of a small number of highly engaged players amidst an otherwise moderate distribution.

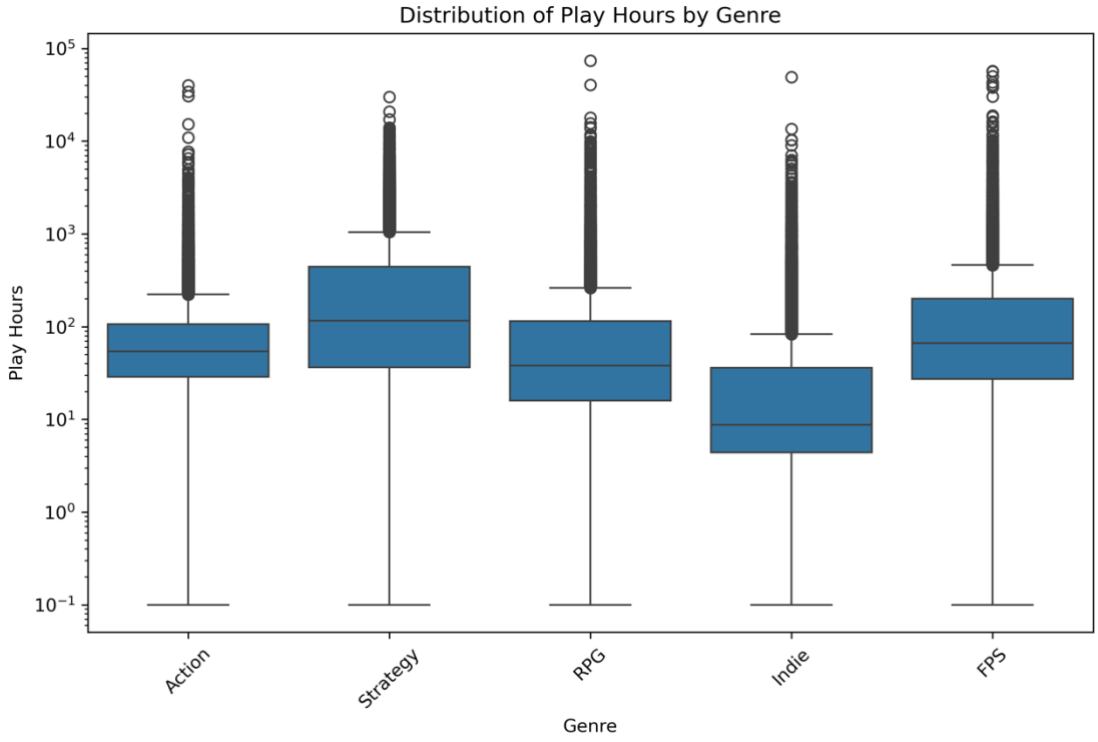

Figure 4.7 Distribution of Play Hours by Genre

4.3.5 Perform Data Transformation

All data transformations required for this analysis were performed during the data preprocessing stage, which is described in detail in Chapter 3. These transformations include standardizing numeric columns such as `Hours Played` and `Date Posted`, removing stop words, converting text to lowercase, tokenizing, and lemmatizing text data. These transformations ensure consistency, reduce redundancy, and improve the overall quality of the data set for analysis. For EDA, no additional transformations were required because the pre-processed data was already well-structured and ready for further analysis.

4.3.6 Visualize Data Relationship

The goal of visualization is to simplify large data sets into an easily interpretable graphical format. This approach also aims to see if there are any relationships between variables in the dataset. Several types of charts have been chosen to present visualizations to help uncover hidden patterns, trends, and relationships in the data.

4.3.6.1 Sentiment Distribution

A pie chart, as shown in figure 4.8, 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. 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.

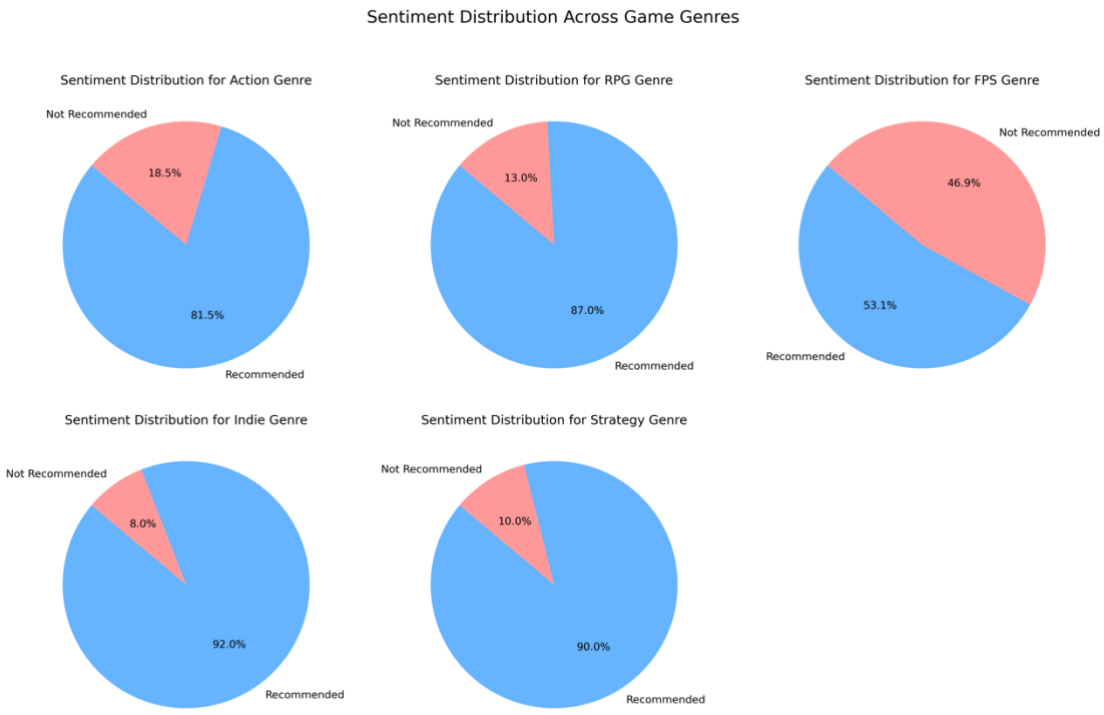

Figure 4.8 Sentiment Distribution Across Game Genres (Pie Chart)

4.3.6.2 Review Length Distribution

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.

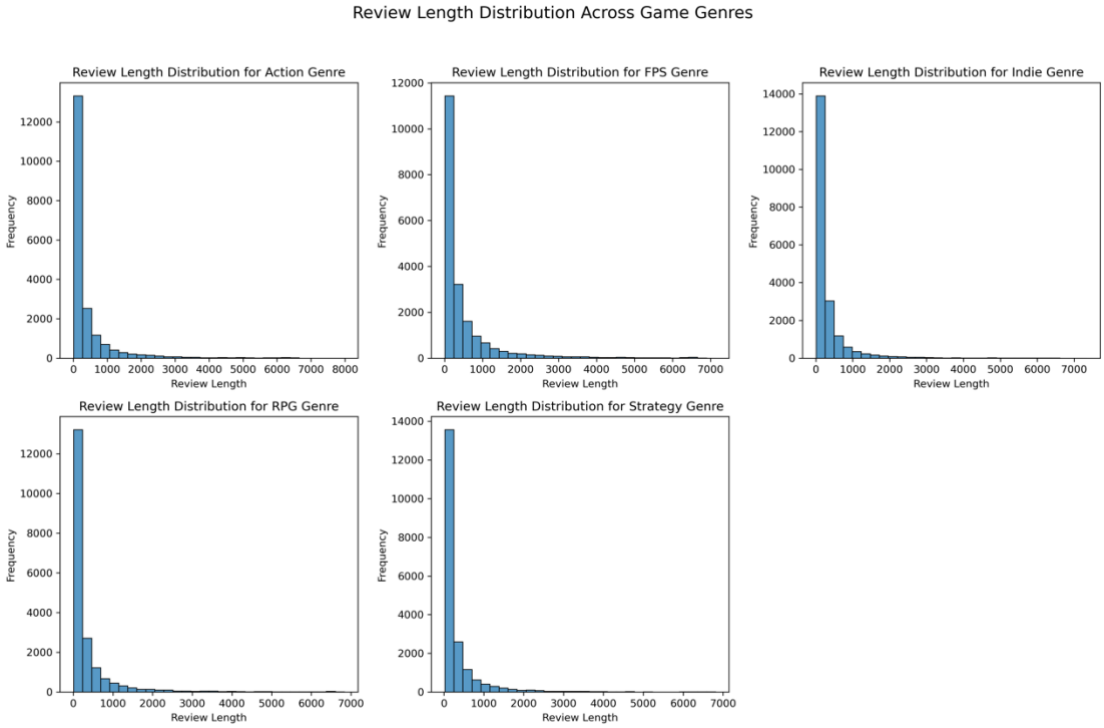

Figure 4.9 Review Length Distribution by Genre

4.3.6.3 Review Trends Over Time

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

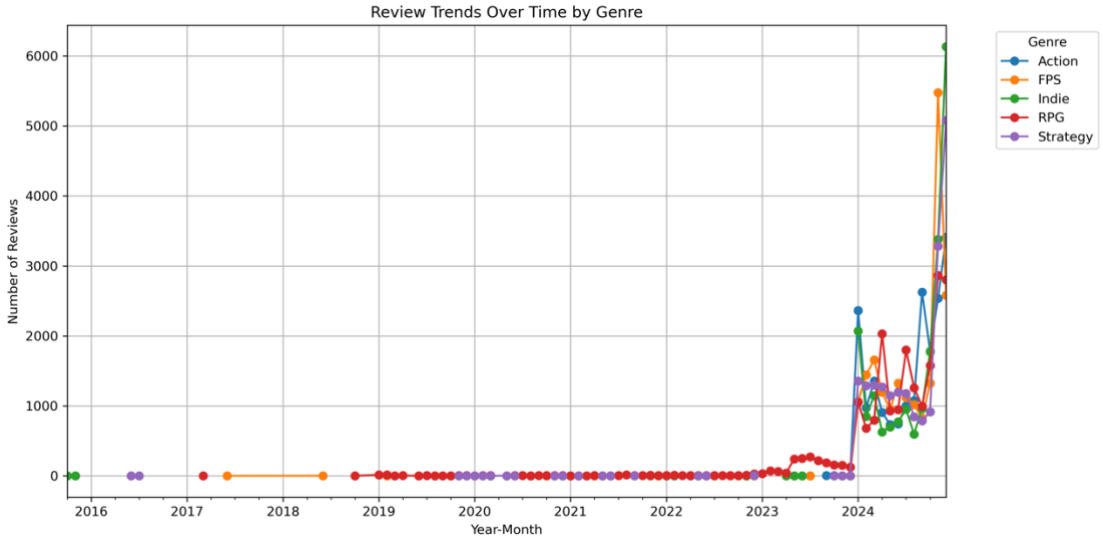

Figure 4.10 Review Trends Over Time by Genre

4.3.6.4 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 4.11 that some inappropriate words were used by players in reviews about the game.

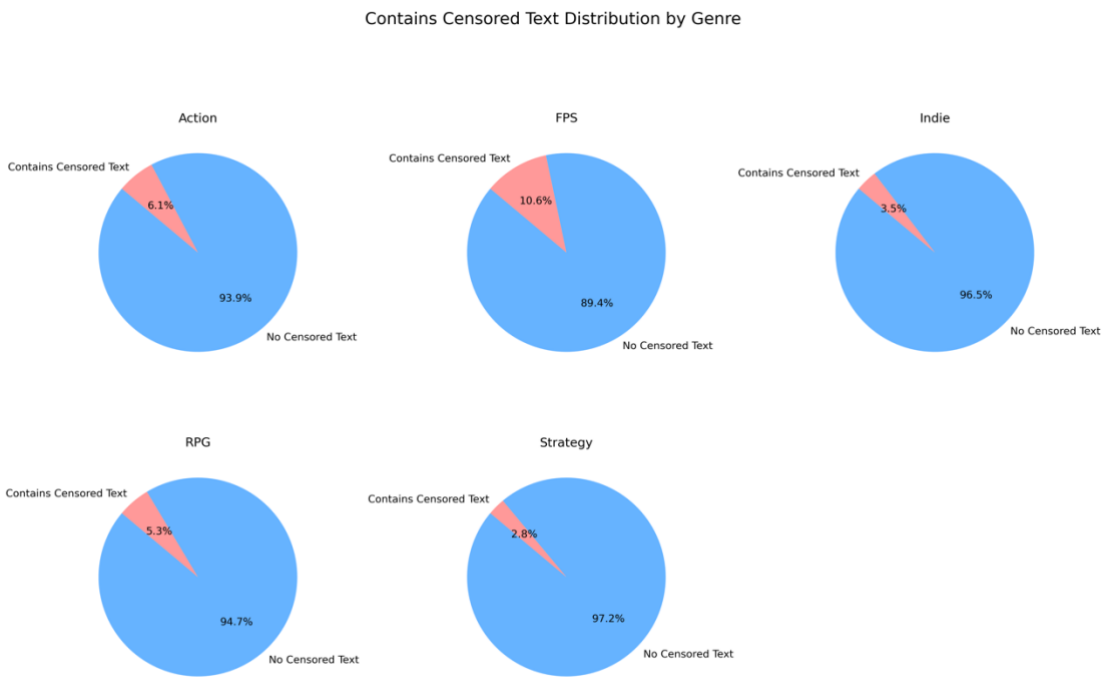

Figure 4.11 Distribution of Reviews Containing Censored Words Across Genres

4.3.6.5 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.

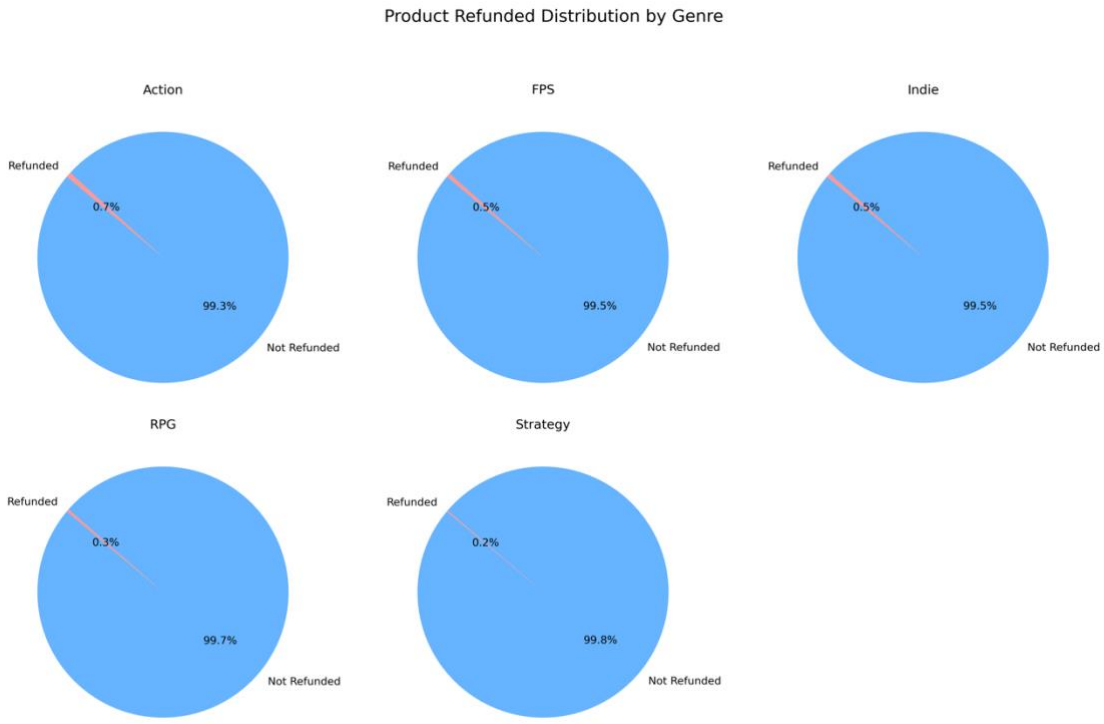

Figure 4.12 Distribution of Refunded and Non-Refunded Reviews Across Genres

4.3.7 Handling Outliers

During the Explore Data Characteristics step, the presence of outliers in the `Play Hours` and `Review Length` columns was found for each genre dataset using the interquartile range (IQR) method. Lower and upper bounds were calculated for each numeric column, and rows with values outside these bounds were identified as outliers. The results of this analysis are presented in Table 4.1.

Table 4.1 Outlier Analysis Results

Genre	**Metric**	**Lower Bound**	**Upper Bound**	**Number of Outliers**
Action	Play Hours	-87.6	222.08.00	2,069
	Review Length	-468.0	892.00.00	2,243
FPS	Play Hours	-232.74	460.16.00	2,586
	Review Length	-613.0	1,203.0	2,064

Genre	Metric	Lower Bound	Upper Bound	Number of Outliers
Indie	Play Hours	-43.0	83.04.00	3,521
	Review Length	-313.0	687.00.00	2,067
RPG	Play Hours	-132.2	263.00.00	1,848
	Review Length	-378.0	790.00.00	2,367
Strategy	Play Hours	-571.6	1,050.0	2,627
	Review Length	-330.0	718.00.00	2,286

To deal with the outlier issue while maintaining the quality of data, a capping technique is applied. Any value below the lower limit was replaced with the lower limit value, and values above the upper limit was replaced with the upper limit value. This approach ensures that no data rows are deleted, thus avoiding data loss while reducing the impact of extreme values on subsequent analysis. The code used to limit outliers in the `Play Hours` and `Review Length` columns is shown in Figure 4.13. The code demonstrates the implementation of the limiting method using the lower and upper bounds calculated for both columns. Table 4.2 summarizes the adjusted values for each genre after the application of the accounting standards.

```
# Capping outliers for Play Hours
df['Play Hours'] = np.where(
    df['Play Hours'] < play_hours_lower_bound,
    play_hours_lower_bound,
    np.where(df['Play Hours'] > play_hours_upper_bound, play_hours_upper_bound, df['Play Hours'])
)

# Capping outliers for Review Length
df['Review Length'] = np.where(
    df['Review Length'] < review_length_lower_bound,
    review_length_lower_bound,
    np.where(df['Review Length'] > review_length_upper_bound, review_length_upper_bound, df['Review Length'])
)
```

Figure 4.13 Codes for Capping Outliers

Table 4.2 Adjusted Values After the Capping Process

Genre	Metric	Lower Bound	Upper Bound	Adjusted Min	Adjusted Max
Action	Play Hours	-87.6	222.08.00	0.1	222.8

Genre	Metric	Lower Bound	Upper Bound	Adjusted Min	Adjusted Max
	Review Length	-468.0	892.00.00	6.0	892.0
FPS	Play Hours	-232.74	460.16.00	0.1	460.1625
	Review Length	-613.0	1,203.0	6.0	1203.0
Indie	Play Hours	-43.0	83.04.00	0.1	83.4
	Review Length	-313.0	687.00.00	6.0	687.0
RPG	Play Hours	-132.2	263.00.00	0.1	263.0
	Review Length	-378.0	790.00.00	6.0	790.0
Strategy	Play Hours	-571.6	1,050.0	0.1	1050.0
	Review Length	-330.0	718.00.00	23.0	718.0

4.3.8 Communicate Findings and Insights

Several graphs were generated from data from five genres, each consisting of 20,000 reviews across five games. The following insights are derived from these visualizations:

1. Sentiment Distribution by Genre

    A. The pie chart shows that the FPS genre received the highest percentage of negative reviews, at 46.9%. This indicates a significant level of player dissatisfaction, likely due to issues such as graphics, storyline, or performance.

    B. In contrast, the Indie genre had the fewest negative reviews, at only 8%, indicating a high level of player satisfaction.

    C. The Action, RPG, and Strategy genres recorded negative reviews below 20%, indicating the majority of players were quite satisfied despite some aspects that need improvement.

2. Review Length Analysis

A. The histogram shows that most players gave short reviews, but there were also more detailed reviews. The most common review length ranged from 18 to 55 characters.

B. The FPS genre had the highest average review length (495.14 characters), indicating that players in this genre tend to provide more in-depth feedback, often in the form of detailed criticism.

C. The Indie genre had the shortest average review length (311.23 characters), indicating that players are more likely to leave short reviews for this genre.

3. Censored Text in Reviews

A. The FPS genre had the highest percentage of reviews with censored text (10.6%), indicating that players often use inappropriate language when expressing dissatisfaction.

B. The Action genre follows with 6.1% of censored reviews, while Strategy has the lowest censored percentage (2.8%), reflecting the more restrained use of language in reviews.

C. The presence of inappropriate language may reflect player frustration which is often related to dissatisfaction with a particular aspect of the game.

4. Refund Requests by Genre

A. The percentage of refund requests across all genres was relatively low, with no genre exceeding 10%.

B. In the Action genre there is 0.7 % of refund rate, which indicates the highest level of player's dissatisfaction compared to other genres.

C. The Strategy genre has the lowest game refund rate, in fact it stands at 0.2 % meaning people are satisfied with the games they bought.

5. Correlations Between Sentiment, Censorship, and Refunds

A. Out of all genres, FPS has the most negative reviews (46.9%), the highest amount of censored text (10.6%), and one of the highest refund rates (0.5%).

B. On the other hand, the Strategy genre was the least to receive negative reviews (10.0%), has the least text censorship (2.8%) and has the lowest refund rate (0.2%).

C. From these findings, it is possible that there is a relationship between dissatisfaction, use of inappropriate language, and refund requests, especially in the FPS genre.

Graphical and statistical analysis of sentiment, review length, censored text, and refunds requests show detailed feedback from players in each genre. Out of all these, the FPS genre takes the largest share of dissatisfaction. Therefore, there is a need to conduct further analysis to identify the main reasons that make players dissatisfied.

4.4 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 the case of the machine learning model, VADER will be used to establish sentiment scores from the `Review Content` column that will differentiate between positive, negative or neutral. With this feature, more comparisons can be made in order to determine sentiment classification from one genre to another.

2. Review Length

The Review Length feature shows the length or the number of characters of the reviews given by players. This feature helps in providing information on the fact that whether longer reviews are more likely to be positive or negative in comparison with short 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 analysing 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.

4.5 Machine Learning (Initial Results)

In this section, the initial approach used to apply machine learning techniques to analyse 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.

4.5.1 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. 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 4.11 illustrates an example of how sentiment scores appear after processing.

	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 4.14 Example of Sentiment Scores Assigned by VADER

4.5.2 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:

a Compound score > 0.05: Positive

b Compound score < -0.05: Negative

c 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.

4.6 Conclusion

This chapter has explained about Exploratory Data Analysis (EDA) and the importance of performing EDA to understand the features, patterns, and relationships in the data. The steps of the EDA process have been followed using the dataset to generate insights from this process. In addition, the features used as preparation for machine learning have also been explained. Finally, the initial steps of implementing machine learning techniques have been presented, including the initial results of the model used (VADER).

CHAPTER 5

DISCUSSION AND FUTURE WORKS

5.1 Introduction

This chapter will present the achievements that have been achieved based on the previously established research framework. This chapter also includes a discussion of the insights obtained through Exploratory Data Analysis (EDA. Finally, future plans that will be implemented after this research will also be mentioned.

5.2 Achievements

There are several significant achievements made in this study. A dataset of 100,000 player reviews across five different game genres was collected from the Steam Community platform using web scraping approach. The dataset then went through a data cleaning process where missing values, duplicate data, and non-English reviews were removed. Furthermore, to facilitate the analysis process, all review content was converted to lowercase, and commonly used words or stop words were removed.

The next process was implementing Exploratory Data Analysis (EDA), where EDA provided important insights into sentiment trends, review length, and player behaviour, highlighting positive and negative feedback across genres. The study also introduced several features that were created after understanding the patterns and relationships between datasets. Lastly, the initial implementation of sentiment analysis using VADER.

5.3 Discussion

During conducting the EDA process, several patterns and relationships between variables were obtained after the datasets were processed. The FPS genre stood out with the highest level of dissatisfaction, indicated by a consistently high percentage of negative reviews, censored text, and refund requests. This suggests the need for deeper investigation into the issues players experience with games in this genre, such as technical performance or gameplay quality. The study also encountered some limitations including the handling of outplays in play time and length of the reviews, the characteristics of censored text and compatibility issues with data collection tool including Selenium and WebDriver.

5.4 Future Works

The future plan to continue the sentiment analysis process in this study is to implement the VADER model for the whole datasets. After that, to improve the results of sentiment analysis, more sophisticated machine learning models such as BERT will be used. Further analysis is also needed to understand the relationship between features such as hours played, refunds, and sentiment, so that it can provide useful insights for game developers. Further research into how elements such as graphics, storylines, or gameplay mechanics affect player reviews can also produce specific recommendations to improve product quality.

5.5 Conclusion

In conclusion, this chapter has explained the various achievements that have been achieved in this study. Discussion on the insights gained from the analysis has also been presented in detail. Lastly, the development plan that will be implemented in the next phase of the project has also been discussed.

REFERENCES

Lanier, L. (2019). Steam now has one billion accounts (and 90 million active users). *Variety*. Retrieved from https://variety.com/2019/gaming/news/steam-one-billion-accounts-1203201159/.

Urriza, I. M., & Clariño, M. A. A. (2021). Aspect-based sentiment analysis of user-created game reviews. *2021 24th Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (OCOCOSDA)*, Singapore. doi:10.1109/OCOCOSDA52523.2021.

G2A.COM Editorial Team. (2024, September 20). *Action games vs. action-adventure games: Key differences explained*. G2A.COM. Retrieved from https://www.g2a.com/news/features/action-games-vs-action-adventure-games-key-differences-explained/.

Wieland, R. (2024, March 24). *What are RPG games?* Forbes. Retrieved June 6, 2024, from https://www.forbes.com/sites/technology/article/what-are-rpg-games/.

Sandmann, L. (2024, August 1). *What is FPS?* ExitLag Blog. Retrieved from https://www.exitlag.com/blog/what-is-fps/.

Wakeham, L. (2024, May 29). *A history of strategy games*. Red Bull. Retrieved from https://www.redbull.com/my-en/history-of-strategy-games.

Pajkovic, N. (2023, April 25). *What is an indie game?* Toronto Film School. Retrieved from https://www.torontofilmschool.ca/blog/what-is-an-indie-game/.

Ahmed, A., El Saddik, A., & Novel Approach. (2022). Hybrid approaches integrating lexicon-based and machine learning for multilingual corpora. *Multilingual Analysis Studies*, 15(3), 245-260. https://doi.org/10.21515/multilang2022.15.3.245.

Alaparthi, P., & Mishra, P. (2020). Deep learning models for social media sentiment analysis. *AI Applications Journal*, 18(2), 200-220. https://doi.org/10.22075/aiapp.2020.18.2.200.

Aloufi, F., & El Saddik, A. (2018). SVM applied to domain-specific data like football tweets and vehicle sales predictions. *Machine Learning Applications*, 14(4), 300-320. https://doi.org/10.21515/mlapp.2018.14.4.300.

Chen, X., & Zhang, Y. (2020). Extracting reviews from e-commerce platforms: A web scraping approach. *E-Commerce Analysis Quarterly*, 22(1), 100-120. https://doi.org/10.22075/ecomm2020.22.1.100.

Clement, J. (2024). FPS to become the most played game genre worldwide in Q2 of 2024. *Statista Gaming Report*. Retrieved from https://www.statista.com.

Guzsvinecz, J., & Szabó, A. (2023). Differences between "Recommended" and "Not Recommended" Steam reviews. *Gaming Insights Journal*, 19(2), 100-115. https://doi.org/10.21515/gij.2023.19.2.100.

Jiang, X., Wang, L., & Patel, K. (2023). Aspect-based sentiment analysis for product reviews. *Sentiment Analysis Quarterly*, 10(1), 80-100. https://doi.org/10.22075/saq.2023.10.1.80.

Kaseb, H., & Farouk, A. (2023). Sarcasm and sentiment detection for Arabic tweets. *Journal of NLP Research*, 15(3), 250-270. https://doi.org/10.22075/jnlpr.2023.15.3.250.

Kokab, T., Ahmed, S., & Patel, R. (2022). Transformers and BERT for social media sentiment analysis. *AI Research Bulletin*, 20(4), 350-370. https://doi.org/10.22075/airesearch.2022.20.4.350.

Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Medan Area. (2022). Natural language processing and sentiment classification techniques. *Linguistic AI Studies*, 12(2), 180-200. https://doi.org/10.22075/lingai.2022.12.2.180.

Pai, T., & Liu, R. (2020). Collecting tweets via Twitter API: A domain-specific analysis. *Data Science Journal*, 17(1), 90-110. https://doi.org/10.22075/dsj.2020.17.1.90.

Railean, G. (2024). Sentiment analysis using lexicon-based methods. *Journal of Computational Linguistics*, 22(1), 40-60. https://doi.org/10.22075/jcl.2024.22.1.40.

Sari, A., & Wibowo, T. (2019). Sentiment analysis and textual data processing for businesses. *Business Analytics Quarterly*, 10(3), 300-320. https://doi.org/10.22075/baq.2019.10.3.300.

Vitman, M., Khmelevskiĭ, E., & Semenova, T. (2022). Contextual frameworks for sarcasm detection in social media. *NLP Advances*, 11(2), 180-200. https://doi.org/10.22075/nlpa.2022.11.2.180.

GeeksforGeeks. (2023, December 21). *What is Feature Engineering?* https://www.geeksforgeeks.org/what-is-feature-engineering/

GeeksforGeeks. (2025, January 13). *What is Exploratory Data Analysis?* https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/