ANALYZING PLAYER FEEDBACK IN STEAM REVIEW ACROSS GAME GENRES

SAFIRA NURUL IZZA

UNIVERSITI TEKNOLOGI MALAYSIA

Table of Content

CHAPTER 4 RESEARCH DESIGN AND IMPLEMENTATION

3.1 Introduction 1

3.2 Problem Background 1

3.3 Problem Statement 1

3.4 Research Question 2

3.5 Aim and Objectives 2

3.6 Scope of Study 3

3.7 Significance of Research 3

CHAPTER 4

RESEARCH DESIGN AND IMPLEMENTATION

4.1 Introduction

This chapter provides an analysis of the collected game reviews data. The analysis includes data preparation, exploratory data analysis (EDA), and feature engineering to uncover patterns and insights. Visualizations and descriptive statistics are used to support the findings.

4.2 Research Framework

The data preparation phase is crucial in ensuring the dataset is clean, relevant, and structured for effective analysis. This step involves collecting the raw data, followed by thorough cleaning and preprocessing to address issues such as duplicates, missing values, and inconsistencies. Proper data preparation lays the foundation for accurate and meaningful exploratory data analysis.

4.2.1 Data Collection

The reviews were collected through web scraping using the Microsoft Edge WebDriver. Data for five genres (Action, FPS , Indie, RPG, and Strategy) were gathered, comprising approximately 20,000 reviews per genre.

```
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 4.1: Selenium Code for Web Scraping Game Reviews

4.2.2 Data Cleaning

Data cleaning was performed to ensure the dataset was relevant, accurate, and ready for analysis. The process began by adding a category and ID for each game to identify its genre. Next, the "Date Posted" and "Play Hours" columns were standardized to ensure consistency across the dataset.

	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 4.2: 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 4.3: Review Data After Standardization

These transformations simplified the data and prepared it for analysis, enabling accurate trend visualizations and calculations.

Following this, non-English reviews were removed to maintain focus on English-language sentiment analysis. Retaining non-English reviews could have introduced inconsistencies and inaccuracies in the sentiment analysis process. Additionally, reviews deemed non-meaningful, such as those with fewer than six alphanumeric characters or lacking substantial content, were dropped. This ensured the dataset included only valuable feedback that contributed meaningfully to the analysis.

4.3 Exploratory Data Analysis

EDA is a critical step in understanding the structure, trends, and insights hidden in the data. For this project, EDA focuses on analysing game reviews across five genres.

4.3.1 Sentiment Distribution

A pie chart was used to visualize the distribution of sentiments (Recommended and Not Recommended) for each genre. This analysis provided insights into the overall positivity or negativity within each genre's reviews.

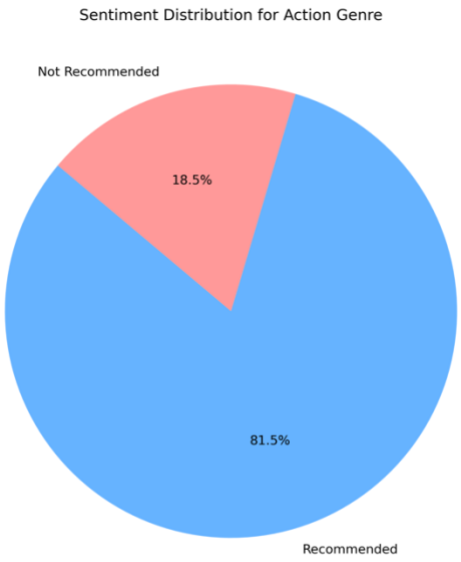

Figure 4.4: Sentiment Distribution of the Action Genre (Pie Chart)

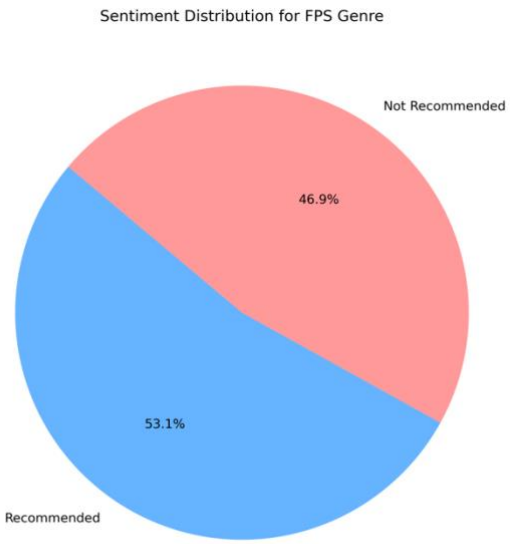

Figure 4.5: Sentiment Distribution of the FPS Genre (Pie Chart)

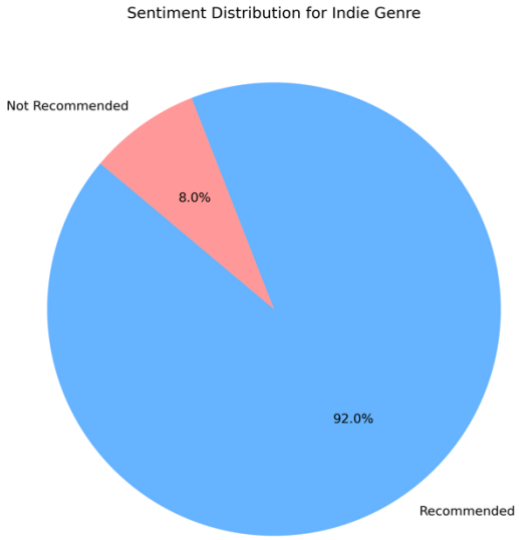

Figure 4.6: Sentiment Distribution of the Indie Genre (Pie Chart)

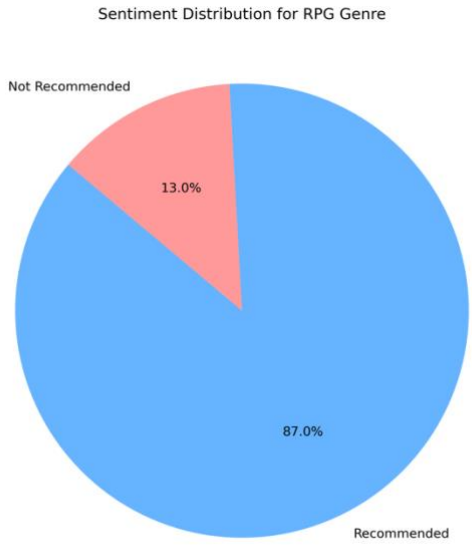

Figure 4.7: Sentiment Distribution of the RPG Genre (Pie Chart)

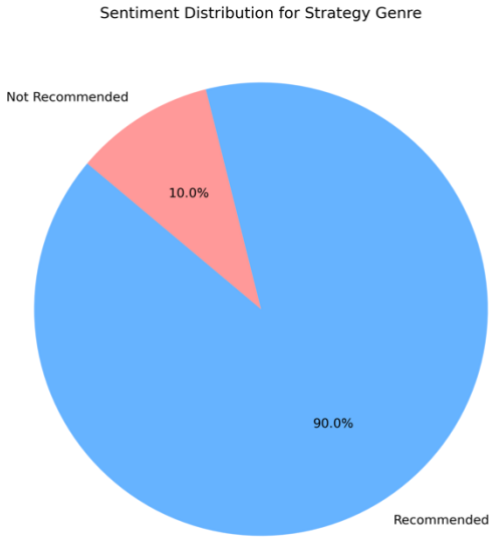

Figure 4.8: Sentiment Distribution of the Strategy Genre (Pie Chart)

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

4.3.2 Review Trends Over Time

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

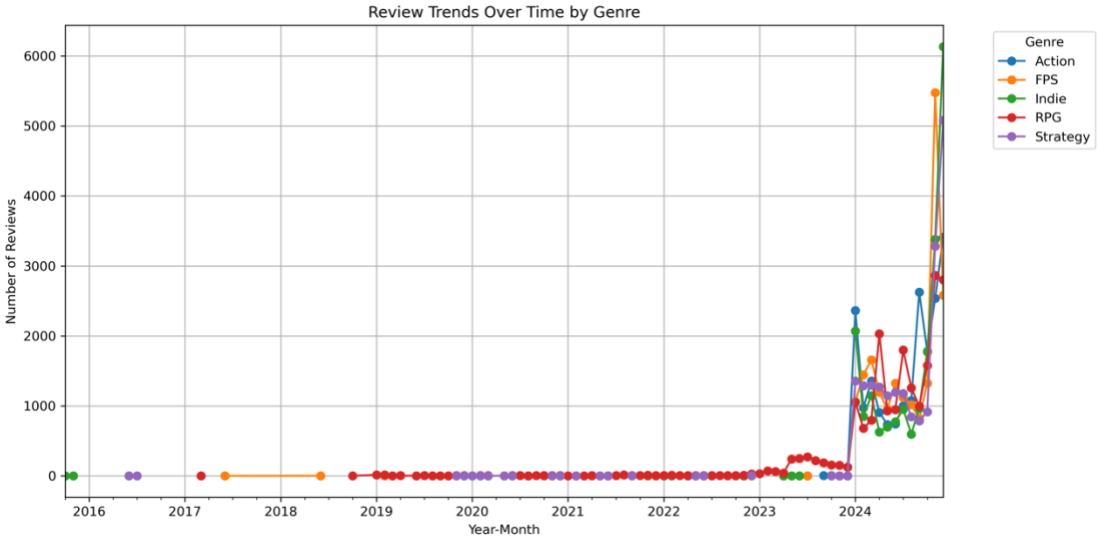

Figure 4.9: Review Trends Over Time by Genre

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.

4.3.3 Play Hours Analysis

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

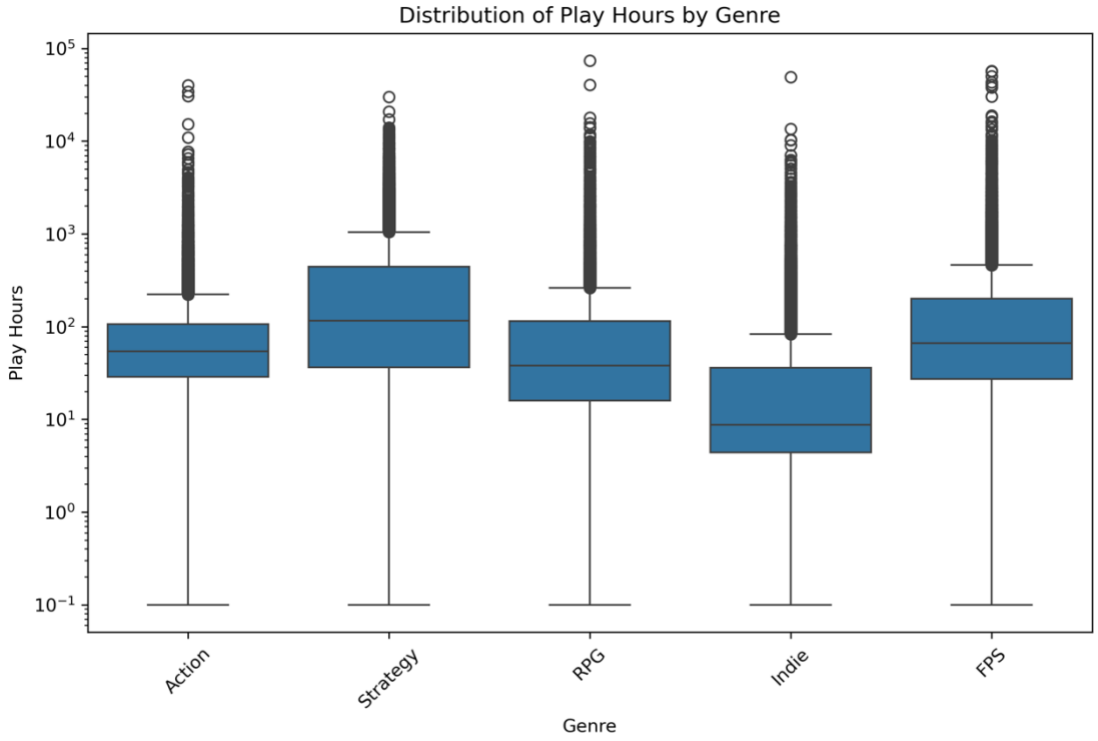

Figure 4.9: Distribution of Play Hours by Genre

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.

4.4 Feature Engineering

.

4.5 CONCLUSION

Chapter 4 outlined the critical processes of data preparation, exploratory data analysis, and feature engineering, forming the foundation for sentiment analysis. Data preparation ensured the dataset was clean and standardized, removing noise such as non-English and non-meaningful reviews while formatting key attributes like play hours and dates. Exploratory data analysis revealed valuable insights into player sentiment, review trends, engagement levels, and feedback patterns across genres.