

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO MULTIDISCIPLINAR

LUCAS DE OLIVEIRA NÓBREGA

**Ciência de Dados para críticas de jogos
eletrônicos**

Prof. Leandro Guimarães Marques Al-
vim, D.Sc.
Orientador

Nova Iguaçu, Março de 2023

Ciência de Dados para críticas de jogos eletrônicos

Lucas de Oliveira Nóbrega

Projeto Final de Curso submetido ao Departamento de Ciência da Computação do Instituto Multidisciplinar da Universidade Federal Rural do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Apresentado por:

Lucas de Oliveira Nóbrega

Aprovado por:

Prof. Leandro Guimarães Marques Alvim, D.Sc.

Prof. Bruno José Dembogurski, D.Sc.

Prof. Fellipe Ribeiro Duarte, D.Sc.

NOVA IGUAÇU, RJ - BRASIL

Março de 2023

Emitido em 13/03/2023

DOCUMENTOS COMPROBATÓRIOS Nº 25/2023 - CoordCGCC (12.28.01.00.00.98)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 19/04/2023 16:04)

BRUNO JOSE DEMBOGURSKI
PROFESSOR DO MAGISTERIO SUPERIOR
DeptCC/IM (12.28.01.00.00.83)
Matrícula: ####249#4

(Assinado digitalmente em 30/03/2023 10:14)

FELLIPE RIBEIRO DUARTE
PROFESSOR DO MAGISTERIO SUPERIOR
DeptCC/IM (12.28.01.00.00.83)
Matrícula: ####890#4

(Assinado digitalmente em 24/03/2023 19:51)

LEANDRO GUIMARAES MARQUES ALVIM
PROFESSOR DO MAGISTERIO SUPERIOR
DeptCC/IM (12.28.01.00.00.83)
Matrícula: ####008#2

(Assinado digitalmente em 26/03/2023 15:16)

LUCAS DE OLIVEIRA NOBREGA
DISCENTE
Matrícula: 2018#####9

Visualize o documento original em <https://sipac.ufrrj.br/documentos/> informando seu número: **25**, ano: **2023**, tipo: **DOCUMENTOS COMPROBATÓRIOS**, data de emissão: **24/03/2023** e o código de verificação: **ff76302ff8**

Agradecimentos

Em primeiro lugar, a Deus, por ter me dado força e saúde para superar as dificuldades e permitir que tudo isso se tornasse possível.

Aos meus pais, Djane e Luiz, por sempre estarem presentes e por todo o esforço investido na minha educação.

Aos meus irmãos, Lidiane e Vitor, pela amizade e por sempre estarem ao meu lado me apoiando.

Aos amigos que fiz durante o curso, André, Eduardo, Marcos Pedro e Rita, por serem os melhores companheiros de trabalho, pela troca de experiências e ajuda mútua.

Aos meus parceiros de time de Maratona de Programação, João Pedro e Thiago, pelo incentivo e troca de aprendizados.

Aos professores, em especial ao meu orientador Leandro Alvim, por todos os conselhos, ajudas e aprendizados.

À Universidade Federal Rural do Rio de Janeiro pela oportunidade de fazer o curso.

RESUMO

Ciência de Dados para críticas de jogos eletrônicos

Lucas de Oliveira Nóbrega

Março/2023

Orientador: Leandro Guimarães Marques Alvim, D.Sc.

O Metacritic é um site que resume as avaliações de filmes, jogos e músicas publicadas por usuários e mídias de entretenimento em uma única nota chamada Metascore. Nos últimos anos, este se tornou um dos parâmetros para o público gamer determinar quais jogos estão em alta ou em baixa e quais produtos ele decide consumir. Muitos trabalhos recentes têm realizado uma análise para entender se o Metascore é uma boa medida para validação de qualidade e valor de um jogo e se existe alguma relação entre essa nota e o número de vendas. Nesse contexto, nota-se a importância de estudar as críticas do Metacritic e, através desses estudos, compreender o motivo pelo alto nível de atenção ao seu conteúdo. Desta forma, o presente trabalho busca contribuir com essa tarefa apresentando uma nova métrica para o cálculo da pontuação geral e uma nova análise das avaliações do Metacritic de forma menos sensível.

ABSTRACT

Ciência de Dados para críticas de jogos eletrônicos

Lucas de Oliveira Nóbrega

Março/2023

Advisor: Leandro Guimarães Marques Alvim, D.Sc.

Metacritic is a website that summarizes reviews of movies, games and music published by users and entertainment media into a single score called Metascore. In recent years, this has become one of the parameters for gamers to determine which games are up or down and which products they decide to consume. Many recent studies have performed an analysis to understand whether the Metascore is a good measure for validating the quality and value of a game and whether there is any relation between this score and the number of sales. In this context, we note the importance of studying Metacritic's reviews and, through these studies, understanding the reason for the high level of attention to its content. Thus, the present work seeks to contribute to this task by presenting a new metric for the calculation of the overall score and a new analysis of Metacritic reviews in a less sensitive way.

Lista de Figuras

Figura 2.1: Representação gráfica do Word Embedding.	13
Figura 2.2: Arquitetura dos modelos CBOW e Skip-gram.	13
Figura 2.3: Cyberpunk 2077 (Playstation 4) no Metacritic.	18
Figura 2.4: The Legend of Zelda: Breath of the Wild (Switch) com o título Must-Play no Metacritic.	18
Figura 2.5: Avaliações do jogo Spore (PC) na Amazon.	20
Figura 2.6: Ellie e Dina de The Last of Us Part II dançando.	20
Figura 3.1: Distribuição de pontuação em avaliações online.	24
Figura 4.1: The Last of Us Part II (Playstation 4) no Metacritic.	31
Figura 4.2: Call of Duty: Modern Warfare (Playstation 4) no Metacritic.	32
Figura 4.3: Elden Ring (PC) no Metacritic.	33
Figura 4.4: Cyberpunk 2077 (PC) no Metacritic.	34
Figura 4.5: Exemplo de Extração de Informações do jogo God of War (PC) no Metacritic.	35
Figura 4.6: Etapa de Web Scraping.	36
Figura 4.7: Etapa de Análise de Sentimentos Orientada a Aspectos.	37
Figura 4.8: Etapa de Geração de Gráficos com Pontuações.	38

Figura 4.9: Etapa de Geração de Gráficos com Termos e Aspectos.	40
Figura 4.10: Histograma de pontuação dos especialistas.	41
Figura 4.11: Histograma de pontuação dos usuários.	41
Figura 4.12: Histograma de pontuação dos usuários na semana de lançamento.	42
Figura 4.13: Top termos negativos dos especialistas.	42
Figura 4.14: Top termos negativos dos usuários.	43
Figura 4.15: Top termos negativos mais frequentes dos usuários.	43
Figura 4.16: Top termos negativos mais frequentes dos usuários na semana de lançamento.	44
Figura 4.17: Top aspectos negativos com maior valor porcentagem vezes polaridade dos especialistas.	44
Figura 4.18: Top aspectos negativos com maior valor porcentagem vezes polaridade dos usuários.	44
Figura 4.19: Top termos positivos dos especialistas.	45
Figura 4.20: Top termos positivos dos usuários.	45
Figura 4.21: Top aspectos positivos com maior valor porcentagem vezes polaridade dos especialistas.	46
Figura 4.22: Top aspectos positivos com maior valor porcentagem vezes polaridade dos usuários.	46
Figura 4.23: Histograma de pontuação dos especialistas.	47
Figura 4.24: Histograma de pontuação dos usuários.	48
Figura 4.25: Top termos negativos dos especialistas.	48
Figura 4.26: Top termos negativos dos usuários.	49

Figura 4.27: Top termos negativos com maior valor porcentagem vezes polaridade dos especialistas.	49
Figura 4.28: Top termos negativos com maior valor porcentagem vezes polaridade dos usuários.	50
Figura 4.29: Top termos positivos dos especialistas.	50
Figura 4.30: Top termos positivos dos usuários.	51
Figura 4.31: Histograma de pontuação dos especialistas.	52
Figura 4.32: Histograma de pontuação dos usuários.	52
Figura 4.33: Top termos negativos dos especialistas.	52
Figura 4.34: Top termos negativos dos usuários.	53
Figura 4.35: Top termos negativos com maior valor porcentagem vezes polaridade dos usuários.	53
Figura 4.36: Top termos positivos com maior valor porcentagem vezes polaridade dos especialistas.	54
Figura 4.37: Top termos positivos com maior valor porcentagem vezes polaridade dos usuários.	55
Figura 4.38: Histograma de pontuação dos especialistas.	56
Figura 4.39: Histograma de pontuação dos usuários.	56
Figura 4.40: Top termos negativos com maior valor porcentagem vezes polaridade dos especialistas.	56
Figura 4.41: Top termos negativos com maior valor porcentagem vezes polaridade dos usuários.	57
Figura 4.42: Top aspectos negativos mais frequentes dos usuários.	57
Figura 4.43: Top termos positivos dos especialistas.	58
Figura 4.44: Top termos positivos dos usuários.	58

Figura A.1: Média de Pontuação com Desvio Padrão ao longo do tempo dos especialistas.	66
Figura A.2: Média de Pontuação com Desvio Padrão ao longo do tempo dos usuários.	67
Figura A.3: Média de Pontuação com Desvio Padrão ao longo do tempo dos usuários na semana de lançamento.	67
Figura A.4: Avaliações por semana dos especialistas.	67
Figura A.5: Avaliações por semana dos usuários.	68
Figura A.6: Avaliações por semana dos usuários na semana de lançamento. . . .	68
Figura A.7: Top termos negativos dos usuários na semana de lançamento. . . .	68
Figura A.8: Top termos negativos mais frequentes dos especialistas.	69
Figura A.9: Top termos negativos com maior valor porcentagem vezes polaridade dos especialistas.	69
Figura A.10:Top termos negativos com maior valor porcentagem vezes polaridade dos usuários.	69
Figura A.11:Top termos negativos com maior valor porcentagem vezes polaridade dos usuários na semana de lançamento.	70
Figura A.12:Top aspectos negativos mais frequentes dos especialistas.	70
Figura A.13:Top aspectos negativos mais frequentes dos usuários.	70
Figura A.14:Top aspectos negativos mais frequentes dos usuários na semana de lançamento.	71
Figura A.15:Top aspectos negativos com maior valor porcentagem vezes polaridade dos usuários na semana de lançamento.	71
Figura A.16:Top termos positivos dos usuários na semana de lançamento. . . .	71
Figura A.17:Top termos positivos mais frequentes dos especialistas.	72

Figura A.18:Top termos positivos mais frequentes dos usuários.	72
Figura A.19:Top termos positivos mais frequentes dos usuários na semana de lançamento.	72
Figura A.20:Top termos positivos com maior valor porcentagem vezes polaridade dos especialistas.	73
Figura A.21:Top termos positivos com maior valor porcentagem vezes polaridade dos usuários.	73
Figura A.22:Top termos positivos com maior valor porcentagem vezes polaridade dos usuários na semana de lançamento.	73
Figura A.23:Top aspectos positivos mais frequentes dos especialistas.	74
Figura A.24:Top aspectos positivos mais frequentes dos usuários.	74
Figura A.25:Top aspectos positivos mais frequentes dos usuários na semana de lançamento.	74
Figura A.26:Top aspectos positivos com maior valor porcentagem vezes polaridade dos usuários na semana de lançamento.	75
 Figura B.1: Média de Pontuação com Desvio Padrão ao longo do tempo dos especialistas.	76
Figura B.2: Média de Pontuação com Desvio Padrão ao longo do tempo dos usuários.	77
Figura B.3: Avaliações por semana dos especialistas.	77
Figura B.4: Avaliações por semana dos usuários.	77
Figura B.5: Top termos negativos mais frequentes dos especialistas.	78
Figura B.6: Top termos negativos mais frequentes dos usuários.	78
Figura B.7: Top aspectos negativos mais frequentes dos especialistas.	78
Figura B.8: Top aspectos negativos mais frequentes dos usuários.	79

Figura B.9: Top aspectos negativos com maior valor porcentagem vezes polaridade dos especialistas.	79
Figura B.10:Top aspectos negativos com maior valor porcentagem vezes polaridade dos usuários.	79
Figura B.11:Top termos positivos mais frequentes dos especialistas.	80
Figura B.12:Top termos positivos mais frequentes dos usuários.	80
Figura B.13:Top termos positivos com maior valor porcentagem vezes polaridade dos especialistas.	80
Figura B.14:Top termos positivos com maior valor porcentagem vezes polaridade dos usuários.	81
Figura B.15:Top aspectos positivos mais frequentes dos especialistas.	81
Figura B.16:Top aspectos positivos mais frequentes dos usuários.	81
Figura B.17:Top aspectos positivos com maior valor porcentagem vezes polaridade dos especialistas.	82
Figura B.18:Top aspectos positivos com maior valor porcentagem vezes polaridade dos usuários.	82
Figura C.1: Média de Pontuação com Desvio Padrão ao longo do tempo dos especialistas.	83
Figura C.2: Média de Pontuação com Desvio Padrão ao longo do tempo dos usuários.	84
Figura C.3: Avaliações por semana dos especialistas.	84
Figura C.4: Avaliações por semana dos usuários.	84
Figura C.5: Top termos negativos mais frequentes dos especialistas.	85
Figura C.6: Top termos negativos mais frequentes dos usuários.	85

Figura C.7: Top termos negativos com maior valor porcentagem vezes polaridade dos especialistas.	85
Figura C.8: Top aspectos negativos mais frequentes dos especialistas.	86
Figura C.9: Top aspectos negativos mais frequentes dos usuários.	86
Figura C.10:Top aspectos negativos com maior valor porcentagem vezes polaridade dos especialistas.	86
Figura C.11:Top aspectos negativos com maior valor porcentagem vezes polaridade dos usuários.	87
Figura C.12:Top termos positivos dos especialistas.	87
Figura C.13:Top termos positivos dos usuários.	87
Figura C.14:Top termos positivos mais frequentes dos especialistas.	88
Figura C.15:Top termos positivos mais frequentes dos usuários.	88
Figura C.16:Top aspectos positivos mais frequentes dos especialistas.	88
Figura C.17:Top aspectos positivos mais frequentes dos usuários.	89
Figura C.18:Top aspectos positivos com maior valor porcentagem vezes polaridade dos especialistas.	89
Figura C.19:Top aspectos positivos com maior valor porcentagem vezes polaridade dos usuários.	89
Figura D.1: Média de Pontuação com Desvio Padrão ao longo do tempo dos especialistas.	90
Figura D.2: Média de Pontuação com Desvio Padrão ao longo do tempo dos usuários.	91
Figura D.3: Avaliações por semana dos especialistas.	91
Figura D.4: Avaliações por semana dos usuários.	91

Figura D.5: Top termos negativos dos especialistas.	92
Figura D.6: Top termos negativos dos usuários.	92
Figura D.7: Top termos negativos mais frequentes dos especialistas.	92
Figura D.8: Top termos negativos mais frequentes dos usuários.	93
Figura D.9: Top aspectos negativos mais frequentes dos especialistas.	93
Figura D.10:Top aspectos negativos com maior valor porcentagem vezes polaridade dos especialistas.	93
Figura D.11:Top aspectos negativos com maior valor porcentagem vezes polaridade dos usuários.	94
Figura D.12:Top termos positivos mais frequentes dos especialistas.	94
Figura D.13:Top termos positivos mais frequentes dos usuários.	94
Figura D.14:Top termos positivos com maior valor porcentagem vezes polaridade dos especialistas.	95
Figura D.15:Top termos positivos com maior valor porcentagem vezes polaridade dos usuários.	95
Figura D.16:Top aspectos positivos mais frequentes dos especialistas.	95
Figura D.17:Top aspectos positivos mais frequentes dos usuários.	96
Figura D.18:Top aspectos positivos com maior valor porcentagem vezes polaridade dos especialistas.	96
Figura D.19:Top aspectos positivos com maior valor porcentagem vezes polaridade dos usuários.	96

Lista de Tabelas

Tabela 1.1: Resumo das pontuações gerais	3
Tabela 2.1: Principais métodos para cálculo de riqueza de vocabulário	15
Tabela 2.2: Riqueza de Vocabulário dos comentários dos usuários do Metacritic NiKCritic e Shadowxxx sobre o jogo Valorant (PC).	16

Lista de Abreviaturas e Siglas

CBOW	Continuous Bag-of-Words
CTTR	Corrected Type-Token Ratio
EA	Electronic Arts
MATTR	Moving Average Type-Token Ratio
MSTTR	Mean Segmental Type-Token Ratio
MTLD	Measure of Lexical Diversity
PLN	Processamento de Linguagem Natural
RTTR	Root Type-Token Ratio
TTR	Type-Token Ratio
WE	Word Embedding

Sumário

Agradecimentos	i
Resumo	ii
Abstract	iii
Lista de Figuras	iv
Lista de Tabelas	xii
Lista de Abreviaturas e Siglas	xiii
1 Introdução	1
1.1 Objetivo	2
1.2 Resumo dos Resultados	2
1.3 Principais Contribuições	3
1.4 Organização do Trabalho	4
2 Fundamentação	5
2.1 Ciência de Dados	5
2.2 Processamento de Linguagem Natural	7

2.2.1	Tokenização	8
2.2.2	Lematização e Stemização	8
2.2.3	Part-of-speech Tagging	9
2.2.4	Análise de Sentimentos Orientada a Aspectos	10
2.2.5	Word Embedding	12
2.2.6	Riqueza de Vocabulário	14
2.3	Web Scraping	16
2.4	Metacritic	17
2.5	Review Bombing	19
3	Proposta	23
3.1	Motivação	23
3.2	Trabalhos Relacionados	26
3.3	Proposta	28
4	Experimentos	30
4.1	Base de Dados	30
4.1.1	The Last of Us Part II (Playstation 4)	30
4.1.2	Call of Duty: Modern Warfare (Playstation 4)	31
4.1.3	Elden Ring (PC)	32
4.1.4	Cyberpunk 2077 (PC)	33
4.2	Metodologia	34
4.2.1	Web Scraping	35
4.2.2	Análise de Sentimento Orientada a Aspectos	36

4.2.3	Geração de Gráficos com Pontuações	37
4.2.4	Geração de Gráficos com Termos e Aspectos	39
4.3	Resultados	39
4.3.1	Experimento 1: The Last of Us Part II (Playstation 4)	40
4.3.2	Experimento 2: Call of Duty: Modern Warfare (Playstation 4)	47
4.3.3	Experimento 3: Elden Ring (PC)	51
4.3.4	Experimento 4: Cyberpunk 2077 (PC)	55
5	Conclusão	60
5.1	Considerações finais	60
5.2	Limitações e trabalhos futuros	61
Referências		63
A	Mais Gráficos e Diagramas gerados no Experimento 1	66
B	Mais Gráficos e Diagramas gerados no Experimento 2	76
C	Mais Gráficos e Diagramas gerados no Experimento 3	83
D	Mais Gráficos e Diagramas gerados no Experimento 4	90

Capítulo 1

Introdução

Com o crescimento da popularidade de jogos eletrônicos, seja como entretenimento ou indústria, cada vez mais pessoas estão envolvidas com vídeo-games. Segundo Clement (2022), o número de *gamers* no ano de 2021 ao redor do mundo é cerca de 3,24 bilhões. Devido a grande quantidade e disponibilidade de opções que existem nos dias atuais, buscar análises de especialistas podem auxiliar o processo de escolha de jogos para esses jogadores.

O Metacritic foi criado com o objetivo de reunir diversas avaliações de filmes, jogos e músicas publicadas por usuários e especialistas. Desta forma, as muitas críticas disponíveis são resumidas de forma que ajude os consumidores a tomar uma decisão sobre como gastar seu tempo e dinheiro em entretenimento. (METACRITICS, 2021)

Nesse contexto, as análises da crítica especializada são simplificadas em uma única nota entre 0 e 100 conhecida como Metascore. Esta é uma média ponderada em que é atribuído maior importância à publicações de melhor qualidade e estatura geral. (METACRITICS, 2021)

Nos últimos anos, houve uma crescente significativa na importância dada às análises e ao Metascore. Segundo Greenwood-Ericksen, Poorman e Papp (2013), pessoas influentes na indústria de jogos não escondem a preocupação com as pontuações atribuídas pelo Metacritic aos jogos com os quais estiverem envolvidos. Isso ocorre

devido a ampla aceitação de que o número de vendas dobra a cada 5 pontos adicionais acima da pontuação 80, ou seja, o Metascore afeta diretamente o número de cópias vendidas.

Desta maneira, percebe-se a necessidade de estudar as análises publicadas pelos usuários e especialistas no Metacritic para que, com isso, possamos analisar e compreender o que se é mais falado de positivo e negativo sobre algum jogo em cada crítica publicada no site.

1.1 Objetivo

O objetivo deste trabalho é contribuir com uma nova análise de dados das avaliações do Metacritic. Para tal, será produzido uma nova métrica para o cálculo da pontuação geral utilizando riqueza de vocabulário. Além disso, através das notas dos usuários e os termos com sentimento positivo e negativo mais utilizados, será realizado uma análise das críticas de uma forma menos sensível a ataques.

1.2 Resumo dos Resultados

Esta seção tem como objetivo apresentar um resumo dos resultados obtidos para a tarefa de análise de dados das avaliações de jogos eletrônicos. Nos experimentos foi utilizado quatro conjuntos de dados extraídos do agregador de avaliações Metacritic. Foi analisado as semelhanças e diferenças das pontuações e conteúdos textuais das críticas de usuários e especialistas. Na Tabela 1.1, estão representados as pontuações gerais para os quatro conjuntos de dados. *The Last of Us Part II* e *Call of Duty: Modern Warfare* são duas bases de dados de jogos que sofreram um bombardeio de críticas negativas de usuários. *Elden Ring* e *Cyberpunk 2077* são duas bases de dados de jogos com várias críticas negativas por conta de problemas no lançamento.

Ao analisar os resultados obtidos nos experimentos usando esses conjuntos de dados, foi observado que a distribuição de pontuação de usuários tendem a ser distorcidas em direção às notas extremas, isto é, 0 ou 10. Por outro lado, os

Tabela 1.1: Resumo das pontuações gerais

	Metacritic		Nova métrica	
	Críticos	Usuários	Críticos	Usuários
The Last of Us Part II	9.3	5.8	9.4	5.4
Call of Duty: Modern Warfare	8.0	3.6	7.8	3.7
Elden Ring	9.4	6.9	9.5	6.5
Cyberpunk 2077	8.6	7.0	8.6	6.4

especialistas possuem uma baixa tendência em atribuir notas negativas, ou seja, entre 0 e 4.

Além disso, é possível observar que os comentários de usuários possuem sentimentos mais extremos e polarizados que os especialistas. Em uma comparação entre os dois, concluimos que os pontos positivos sobre um jogo eletrônico são similares. No entanto, os utilizadores do Metacritic encontram diversos outros pontos negativos que não são comentados pela mídia especializada.

1.3 Principais Contribuições

Nesta seção, será apresentado de maneira resumida as principais contribuições do presente trabalho em termos do estudo do problema, proposta e disponibilização do código fonte.

- a) Estudo sobre as avaliações de jogos eletrônicos do Metacritic

Neste trabalho, foi produzido um nova análise das avaliações de usuários e especialistas do Metacritic sobre jogos eletrônicos. O estudo torna-se ainda mais importante devido a falta de pesquisas sobre o conteúdo textual dessas críticas.

- b) Nova métrica para cálculo de pontuação geral

Uma segunda importante contribuição obtida com este trabalho encontra-se na nova métrica para o cálculo da pontuação geral. Este torna-se significativo, já que o agregador de avaliações Metacritic não disponibiliza como é calculado o Metascore e a nota de usuário.

- c) Comparação entre as avaliações de usuários e especialistas

Outra considerável contribuição é a comparação entre as avaliações de usuários e especialistas realizada nos experimentos deste trabalho. Novamente, este é um importante estudo devido a escassez de pesquisas sobre o conteúdo textual das críticas do Metacritic.

d) Código fonte

Por fim, uma outra importante contribuição deste trabalho é o código fonte¹ que está disponível em um repositório público para que possa ser usado em uma pesquisa futura.

1.4 Organização do Trabalho

No Capítulo 2 será apresentado todos os principais conceitos utilizados para que o leitor consiga ter a capacidade de entender o presente trabalho. O Capítulo 3 explicará como será a solução, exibindo a proposta utilizada para alcançar o objetivo principal. O Capítulo 4 fornecerá como o projeto foi desenvolvido, revelando quais as ferramentas foram utilizadas para implementação do sistema e os seus motivos. Além disso, será mostrado os experimentos realizados e uma análise dos resultados obtidos. Por fim, o Capítulo 5 tratará sobre as conclusões, considerações finais, limitações encontradas e possíveis trabalhos futuros.

¹<<https://github.com/LucasONobrega/Aspectos-Metacritic-TCC>>

Capítulo 2

Fundamentação

Neste capítulo, será fornecido os principais conceitos teóricos necessários para que a proposta do trabalho seja entendida. Os seguintes assuntos serão apresentados: Ciência de Dados, Processamento de Linguagem Natural, *Web Scraping*, Metacritic e *Review Bombing*.

2.1 Ciência de Dados

Ciência de Dados é a combinação de ciência da computação, matemática, estatística e conhecimento de negócio para estudar e extrair informações importantes dos dados. Assim, seu objetivo é criar técnicas, tecnologias e teorias para analisar os dados extraídos de fontes diversas, gerar algum significado e transformá-los em conhecimento a fim de auxiliar em tomadas de decisão. (AWS, 2023)

Em 2008, o economista-chefe do Google e professor da UC Berkeley Hal Varian definiu cientista de dados como aquele capaz de coletar dados, entendê-los, processá-los, extrair valor, visualizá-los e comunicá-los (BERKELEY, 2023). Isto significa que um cientista de dados deve conhecer e entender todo o ciclo de vida da Ciência de Dados: Obter, Manter, Explorar, Analisar e Interpretar (AWS, 2023).

- a) Obter: Esta etapa consiste na aquisição de dados - estruturados e não estruturados - a partir de uma ou mais fontes utilizando algum método de captura

como web scraping, streaming de dados em tempo real ou até mesmo entrada manual.

- b) Manter: Após obter os dados, é necessário guardá-los em algum sistema de armazenamento. Em seguida, esta etapa consiste em realizar o processo de limpeza de dados. Assim, os dados são padronizados em algum formato específico e problemas como ausência ou informações atípicas são tratadas.
- c) Explorar: A etapa de exploração de dados é uma análise de dados preliminar usada para verificar se existe algum padrão que pode ser útil. Para isso, é utilizado estatísticas descritivas e ferramentas de visualização.
- d) Analisar: Na etapa de análise dos dados é utilizado algoritmos e técnicas de aprendizado de máquina para obter informações mais detalhadas, prever resultados, classificar ou agrupar.
- e) Interpretar: Por fim, a última etapa consiste na visualização dos dados para facilitar a compreensão das informações obtidas. Para isso, é utilizado diagramas, gráficos e tabelas.

Com isto, nota-se que a Ciência de Dados pode ser utilizada para estudar os dados de diversas maneiras. Os quatro principais tipos de análise de dados são: Análise descritiva, Análise diagnóstica, Análise preditiva e Análise prescritiva. (AWS, 2023)

- a) Análise descritiva: Este tipo de análise obtém informações sobre o que aconteceu ou o que está acontecendo com os dados. Para tal, é utilizado alguma forma de visualização de dados, como diagramas, gráficos ou tabelas. A título de exemplo, uma empresa de reserva de voos pode registrar quantas passagens aéreas são reservadas a cada dia. Desta maneira, será possível analisar quando ocorre picos, quedas e os meses mais lucrativos para este serviço.
- b) Análise diagnóstica: Este tipo de análise obtém informações mais detalhadas sobre os dados para entender o motivo de alguma coisa ter acontecido. Para isso, técnicas de detalhamento, descoberta de dados, mineração de dados e correlações são utilizadas de forma que padrões possam ser encontrados.

Como exemplo, no mesmo cenário da empresa de reserva de voos, é possível realizar um detalhamento dos dados para descobrir o motivo do pico de venda em algum mês específico.

- c) Análise preditiva: Este tipo de análise utiliza dados históricos para fazer previsões precisas sobre algo que pode ocorrer no futuro. Com esse propósito, técnicas de aprendizado de máquina, previsão e modelagem preditiva podem ser utilizadas. Por exemplo, continuando no cenário da empresa de reserva de voos, é possível analisar dados dos anos anteriores para prever padrões de reservas de voos para um determinado destino e mês.
- d) Análise prescritiva: Este tipo de análise também utiliza dados históricos para, além de prever como na Análise preditiva, recomendar uma resposta ideal para esse resultado. Para isso, são utilizados técnicas de simulação, redes neurais e aprendizado de máquina para que seja sugerido o melhor plano de ação possível. Como exemplo, no cenário da empresa de reserva de voos, é possível analisar os dados dos anos anteriores para maximizar o lucro em um possível próximo pico de vendas.

2.2 Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) é uma área da computação e inteligência artificial (JURAFSKY; MARTIN, 2010) que consiste em estudar e desenvolver modelos computacionais para extrair representações e significados mais completos de textos livres escritos em linguagem natural (INDURKHYA; DAMERAU, 2010).

Dessa maneira, segundo Liddy (2001), o PLN permite aos computadores analisar e representar textos que ocorrem naturalmente em um ou mais níveis de análise linguística. Estes níveis são:

- a) Fonológico: interpretar os sons da fala dentro e entre as palavras.
- b) Morfológico: analisar os componentes das palavras, como prefixo, sufixo e raíz.

- c) Léxico: analisar significado lexical e classe gramatical de uma palavra.
- d) Sintático: analisar palavras em uma frase para descobrir a estrutura gramatical da frase.
- e) Semântico: determinar o possível significado de uma frase, incluindo a desambiguação de palavras em contexto.
- f) Discurso: interpretar a estrutura e o significado transmitido por textos maiores que uma frase.
- g) Pragmático: compreender o uso intencional da linguagem em situações, especialmente os aspectos da linguagem que requerem conhecimento do mundo.

Os níveis de análise linguística descritos anteriormente muitas vezes podem ser divididos em tarefas menores e mais simples (PINTO, 2015). Em seguida, serão apresentados algumas dessas tarefas básicas, como Tokenização, Lematização, Stemização e *Part-of-speech Tagging*.

2.2.1 Tokenização

A Tokenização, ou segmentação de palavras, é uma das tarefas mais simples de normalização de texto e a primeira etapa de análise em PLN. Seu objetivo é dividir um texto em uma sequência de tokens, localizando os limites das palavras, as partes em que uma palavra termina e outra começa (INDURKHYA; DAMERAU, 2010).

Por exemplo, a frase:

“Este é um exemplo.”

é dividida no seguinte conjunto de tokens:

[Este] [é] [um] [exemplo] [.].

2.2.2 Lematização e Stemização

Lematização e Stemização é uma outra parte importante da normalização de texto em que o objetivo é relacionar as variantes morfológicas às formas canônicas

das palavras, ou seja, como as palavras aparecem em um dicionário (INDURKHYA; DAMERAU, 2010).

A Lematização consiste em determinar o lema de uma palavra. Um dos métodos envolve análise morfológica completa da palavra, ou seja, a partir de qual *stem* e afixos a palavra é construída (JURAFSKY; MARTIN, 2010). Para isso, é preciso utilizar um dicionário de lemas que contém informações semânticas e sintáticas invariantes (INDURKHYA; DAMERAU, 2010). Deste modo, se torna possível saber se duas palavras têm a mesma raiz, apesar de suas diferenças superficiais (JURAFSKY; MARTIN, 2010). Por exemplo, a frase:

“Estou aprendendo processamento de linguagem natural.”

consiste no seguinte conjunto de lemas:

[estar] [aprender] [processar] [de] [língua] [natural].

Stemização é uma versão mais simples da Lematização em que removemos somente o sufixo do final de uma palavra (JURAFSKY; MARTIN, 2010). Sendo assim, diferente da Lematização, não existe uma garantia de que o *stem* seja uma palavra gramaticalmente correta. Por exemplo, a mesma frase anterior equivale ao seguinte conjunto de *stems*:

[est] [aprend] [process] [de] [lingu] [natur]

2.2.3 Part-of-speech Tagging

Part-of-speech Tagging, ou *POS Tagging*, é o processo de atribuir uma classe gramatical a cada palavra de um texto (JURAFSKY; MARTIN, 2010). Por exemplo, a frase:

“O carro de João é cinza”

possui a seguinte classificação:

[Artigo] [Substantivo] [Ad posição] [Substantivo] [Verbo] [Adjetivo].

Esta é uma tarefa não trivial em virtude de que algumas palavras possuem mais de uma classe gramatical dependendo da frase em que é usada. Portanto, o objetivo do POS Tagging é resolver essas ambiguidades, escolhendo a *tag* adequada para o contexto (JURAFSKY; MARTIN, 2010).

As técnicas que foram apresentadas são frequentemente utilizadas como base em atividades mais complexas de PLN. A seguir, serão apresentados algumas dessas tarefas, tais como: Análise de Sentimentos Orientada a Aspectos, *Word Embedding* e Riqueza de Vocabulário.

2.2.4 Análise de Sentimentos Orientada a Aspectos

A Análise de Sentimento, ou Mineração de Opinião, é a tarefa de PLN responsável por definir técnicas computacionais que extraem opinião, sentimento e subjetividade de textos em linguagem natural (PANG; LEE, 2008).

Segundo Filho (2017), as informações que podem ser extraídas de um texto são definidas da seguinte forma:

- a) Sentimento: termo genérico para designar todo texto que expressa características positivas, negativas ou neutras.
- b) Subjetividade: consiste em um texto que possui sentimentos, pontos de vista ou crenças pessoais.
- c) Opinião: é representado por dois componentes: um aspecto e um sentimento associado.

Um dos principais tópicos de pesquisa da Análise de Sentimentos é a classificação de sentimentos. Esta pode ser usada para as diferentes tarefas e problemas, dependendo do seu objetivo. As mais utilizadas são: (SENTIMENT...,)

- a) Identificação de polaridade: classifica a polaridade de um texto opinativo como uma opinião positiva, negativa ou neutra.

- b) Detecção de emoções: identifica os sentimentos e emoções de um texto, como irritação, felicidade e tristeza.
- c) Detecção de urgência: identifica se o texto expressa algum tipo de urgência ou não.
- d) Deteção de intenções: identifica se o texto expressa algum tipo de interesse ou não.

Apesar disso, a análise de sentimentos se concentra principalmente na identificação de polaridade. Esta tarefa pode ser realizado em três níveis diferentes: documento, frase e aspecto (INDURKHYA; DAMERAU, 2010; SENTIMENT...,).

No nível do documento, a classificação de sentimentos é feita com base na análise do texto como um todo. Neste caso, é considerado que no documento exista apenas um tópico que possui certa polaridade. Além disso, geralmente, o sentimento é somente positivo ou negativo. (INDURKHYA; DAMERAU, 2010; FILHO, 2017)

Em relação ao nível da frase, é classificado a polaridade de cada sentença do texto. Além das opções positivo e negativo, em alguns casos, também é utilizado uma terceira possibilidade: neutro. Este é usado para indicar a falta de sentimento ou um sentimento indeterminado. (FILHO, 2017)

O nível de análise que este trabalho utiliza e o foco desta subseção é o de aspectos. Neste nível, determina-se a polaridade do sentimento expresso em características, atributos ou entidades presentes no texto. O fato de que um sentimento sempre é expresso junto de um aspecto é uma das motivações deste tipo de análise de sentimentos. Dessa forma, ao invés de considerar construções textuais, olha-se diretamente para a opinião existente no texto. (FILHO, 2017)

A Análise de Sentimentos Orientada a Aspectos é composta por três principais tarefas, são elas: Extração de Aspectos, Identificação de Sinônimos e Classificação de Sentimentos. (INDURKHYA; DAMERAU, 2010)

A tarefa de Extração de Aspectos é responsável por identificar as características de um objeto e seus modificadores (INDURKHYA; DAMERAU, 2010). Por exemplo, na frase:

“A história desse filme é emocionante”,

a característica do objeto extraída é “emocionante história”, sendo que o aspecto é a palavra “história” e o modificador é “emocionante”.

A tarefa Identificação de Sinônimos consiste em agrupar entidades, aspectos e modificadores (FILHO, 2017). Esta é uma etapa importante, pois assim é possível identificar que, por exemplo, os aspectos “celular” e “telefone móvel” ou os modificadores “excelente” e “maravilhoso” significam a mesma coisa.

Por fim, após a Extração de Aspectos e Identificação de Sinônimos, precisamos realizar a Classificação de Sentimentos, ou seja, identificar se a polaridade do aspecto é positiva ou negativa de acordo com os seus modificadores (FILHO, 2017). Por exemplo, a característica do objeto “história emocionante” é positivo.

2.2.5 Word Embedding

Word Embedding (WE) é uma forma de representação de texto em que o objetivo é transformar uma palavra em um ponto de um espaço semântico multidimensional (JURAFSKY; MARTIN, 2010). Isto significa que o WE permite representar qualquer palavra de um dicionário em um vetor com dimensão reduzida (PINTO, 2015).

Desta maneira, torna-se necessário que o WE permita que quaisquer duas palavras que são usadas em contexto semântico e sintático similares possuam vetores de representação semelhantes (PINTO, 2015). Por exemplo, como é demonstrado na Figura 2.1, palavras como “rei” e “rainha” ou “homem” e “mulher” devem possuir vetores parecidos. Além disso, uma operação vetorial como “rei” - “homem” + “mulher” deve resultar em um vetor próximo de “rainha”.

Com este intuito, diversos modelos foram propostos para estimar a transformação para o espaço vetorial. Existem duas abordagens para aprender representações de palavras com eficiência em grandes quantidades de dados de texto, são eles: Continuous Bag-of-Words (CBOW) e Skip-gram. No modelo CBOW, uma janela pré-definida de palavras vizinhas e seu vetor de contexto são utilizados para prever a palavra alvo. Por outro lado, o modelo Skip-gram aprende a prever as palavras

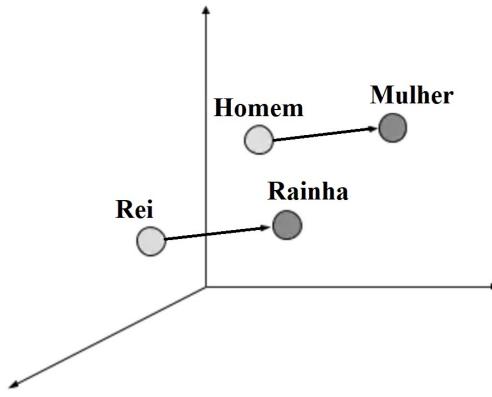


Figura 2.1: Representação gráfica do Word Embedding.

Adaptado de: <<https://medium.com/luizalabs/similaridade-entre-t%C3%A9cnicas-de-produtos-com-word2vec-5e26199862f0>>. Acesso em: 15 dez. 2022

vizinhas com base na palavra atual. As arquiteturas desses dois modelos são mostradas na Figura 2.2. Na prática, o Skip-gram oferece melhores representações de palavras e o CBOW é mais rápido e adequado para conjuntos de dados maiores. (MIKOLOV et al., 2013; MIKOLOV; LE; SUTSKEVER, 2013)

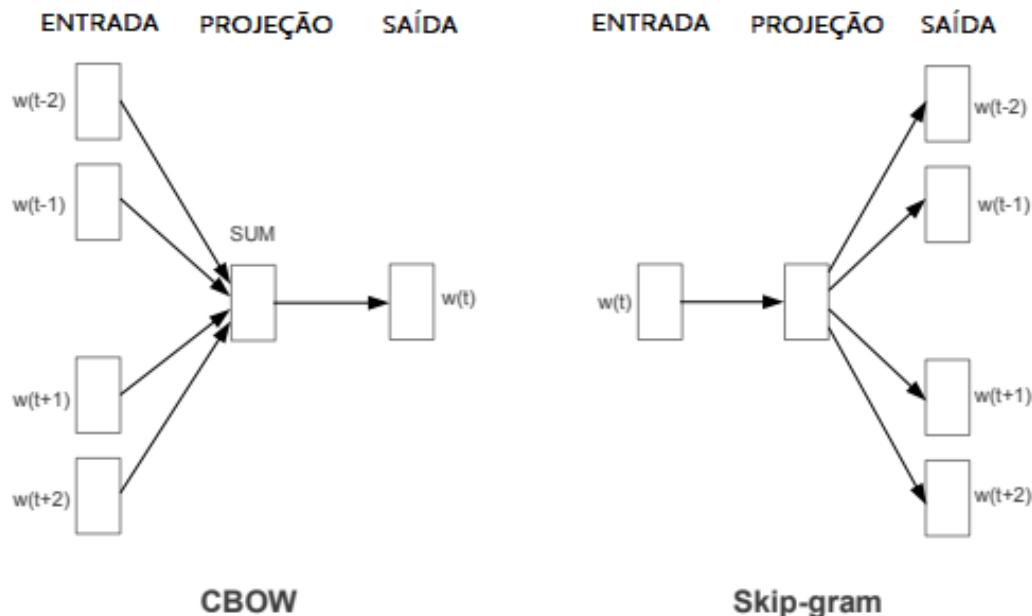


Figura 2.2: Arquitetura dos modelos CBOW e Skip-gram.

Adaptado de: Mikolov et al. (2013). Acesso em: 16 dez. 2022

Umas das possíveis e mais utilizadas aplicações do WE é a similaridade de palavras. Neste caso, podemos analisar o quanto duas palavras são semelhantes entre si. Para isso, é necessário calcular a distância entre as representações vetoriais dos termos. Quanto menor for esse valor, mais similares entre si são as palavras.

2.2.6 Riqueza de Vocabulário

Riqueza de Vocabulário é definida como a diversidade lexical nas palavras utilizadas em um texto por um falante ou escritor. Desta forma, torna-se possível analisar a sua capacidade em acessar e recuperar os termos de uma base de conhecimento para a construção de unidades linguísticas mais complexas. (FERGADIOTIS; WRIGHT; WESTA, 2013)

Em PLN, podemos determinar Riqueza de Vocabulário como uma métrica que estabelece a relação entre o número de termos únicos t e o número total de palavras w . A Tabela 2.1 representa as principais e mais utilizadas medidas, Type-Token Ratio (TTR), Root Type-Token Ratio (RTTR), Corrected Type-Token Ratio (CTTR), Herdan, Summer, Dugast, Maas, Mean Segmental Type-Token Ratio (MSTTR), Moving Average Type-Token Ratio (MATTR) e Measure of Lexical Diversity (MTLD), e como elas são calculadas.

O TTR é uma das abordagens mais utilizadas para medir a riqueza de vocabulário de um texto e é calculada com seguinte fórmula: $TTR = t/w$. Entretanto, por variar em função do número de palavras, essa é uma medida que é falha. Quanto maior o tamanho do texto, é mais provável de existir palavras repetidas. Sendo assim, amostras mais curtas geralmente parecem ser mais ricas. (FERGADIOTIS; WRIGHT; WESTA, 2013)

Por este motivo, surgiram novas técnicas a partir de correções simples no TTR, como RTTR e CTTR (TORRUELLA; CAPSADA, 2013). Essas medidas são calculadas com as seguintes fórmulas: $RTTR = t/\sqrt{w}$ e $CTTR = t/\sqrt{2w}$.

Além disso, mais recentemente foi desenvolvido técnicas que usam fórmulas baseadas em funções logarítmicas, como Herdan, Summer, Dugast e Maas. Segundo

Tabela 2.1: Principais métodos para cálculo de riqueza de vocabulário

Método	Cálculo
TTR	t/w
RTTR	t/\sqrt{w}
CTTR	$t/\sqrt{2w}$
Herdan	$\log t / \log w$
Summer	$\log(\log t) / \log(\log w)$
Dugast	$(\log w)^2 / (\log w - \log t)$
Maas	$(\log w - \log t) / (\log w)^2$
MSTTR	Divide os tokens em segmentos do tamanho dado, TTR para cada segmento é calculado e a média desses valores é retornada.
MATTR	Calcula TTRs para uma janela móvel de tokens do primeiro ao último token, calculando um TTR para cada janela. O MATTR é a média dos TTRs.
MTLD	Calcula o comprimento médio das sequências de tokens sequenciais em um texto que mantém um determinado valor TTR.

Torruella e Capsada (2013), essas abordagens tendem a se adaptar melhor ao comportamento da relação entre o número de termos únicos e o número total de palavras. Essas medidas são calculadas com as seguintes fórmulas: $Herdan = \log(t)/\log(w)$, $Summer = \log(\log(t))/\log(\log(w))$, $Dugast = \log^2(w)/(\log(w) - \log(t))$ e $Maas = (\log(w) - \log(t))/\log^2(w)$.

Por fim, existe uma última classe de abordagens para medir riqueza de vocabulário em que o valor é obtido através de cálculos mais complexos, são eles: MSTTR, MATTR e MTLD. Em MSTTR, o texto é dividido em segmentos com o mesmo número de palavras. Já no MATTR, é utilizado uma janela móvel de tokens, ou seja, o texto é dividido em subtextos sobrepostos de mesmo comprimento. Para cada uma dessas partes é calculado o valor do TTR e o valor final é obtido usando média aritmética. O MTLD é semelhante ao MSTTR, porém o comprimento dos segmentos pode variar dependendo do TTR obtido na janela anterior. (FERGADIOTIS; WRIGHT; WESTA, 2013; TORRUELLA; CAPSADA, 2013)

Como forma de exemplo será utilizado dois comentários extraídos do Metacritic com tradução livre sobre o jogo “VALORANT” (PC), sendo um texto rico e outro pobre de vocabulários. O usuário NiKCritic escreveu:

“O jogo é meu FPS favorito. Muita gente fala que os gráficos são ruins, mas não é verdade. Tem gráficos bons e as expressões dos personagens são perceptíveis. Algo que gosto e o que o diferencia dos outros FPS é que cada personagem tem habilidades únicas e cada uma tem a ver com o seu design. Por isso, me parece um FPS incrível que difere dos demais e recomendo muito”.

Já o usuário Shadowxxx comentou o seguinte:

“nojento lixo nojento lixo nojento lixo nojento lixo nojento”.

A Tabela 2.2 mostra a riqueza de vocabulário desses comentários utilizando as métricas apresentadas.

Tabela 2.2: Riqueza de Vocabulário dos comentários dos usuários do Metacritic NiKCritic e Shadowxxx sobre o jogo Valorant (PC).

	NiKCritic	Shadowxxx
TTR	0.70	0.22
RTTR	5.85	0.66
CTTR	4.14	0.47
Herdan	0.91	0.31
Summer	0.93	-0.46
Dugast	50.60	3.20
Maas	0.01	0.31
MSTTR	0.69	0.25
MATTR	0.70	0.25
MTLD	60.55	3.00

2.3 Web Scraping

Atualmente, um grande volume de dados é gerado todos os dias por diversas e diferentes fontes na internet (OUSSOUS et al., 2018). Devido a toda essa criação de dados, a *International Data Corporation* prevê que a esfera de dados global crescerá de 33 *Zettabytes* em 2018 para 175 *Zettabytes* em 2025 (REINSEL; GANTZ; RYDNING, 2018).

Desta forma, por conta da grande quantidade e variedade dos dados, torna-se necessário coletar e organizar essas informações. Este processo dificilmente pode ser

feito de forma manual com o tradicional copia e cola. Por isso, para automatizar a coleta de dados da *web*, surgiu a técnica conhecida como *Web Scraping*. (SIRISURIYA, 2015)

O *Web Scraping*, ou Extração de Dados da *Web*, é definido como o processo de extração automática de informações úteis de páginas da internet (DASTIDAR; BANERJEE; SENGUPTA, 2016). O objetivo principal da técnica é extrair dados não estruturados de sites e os transformar em dados estruturados que podem ser armazenados e analisados em um banco de dados (SIRISURIYA, 2015).

2.4 Metacritic

O Metacritic é um site criado em 2001 por Jason Dietz, Marc Doyle e Julie Roberts com o objetivo de reunir as muitas críticas de filmes, jogos e músicas disponíveis (METACRITICS, 2021). Desta forma, em virtude da grande quantidade e disponibilidade de opções de entretenimento, o Metacritic consegue auxiliar os seus usuários a tomar uma decisão sobre o que consumir.

Com esse intuito, as análises da crítica especializada são resumidas em uma única nota geral conhecida como Metascore. Este varia de 0 a 100, sendo que pontuações mais altas indicam melhores avaliações. Além disso, o Metascore é destacado com uma de três cores: verde para avaliações favoráveis, amarelo para avaliações mistas e vermelho para avaliações desfavoráveis. Para o cálculo do Metascore, usa-se uma média ponderada em que publicações de melhor qualidade e estatura geral possuem um peso maior. (METACRITICS, 2021)

Além disso, de forma a ajudar os usuários a escolher rapidamente o que há de melhor em entretenimento, o Metacritic designa os títulos “*Must-See*” (filmes), “*Must-Watch*” (séries), “*Must-Play*” (jogos) para os mais bem avaliados pela mídia especializada. Para receber tal designação, é preciso possuir pelo menos 15 avaliações e atingir um Metascore mínimo de 81 para filmes/séries e 90 para jogos. Como exemplo, os filmes “*The Godfather*” e “*Toy Story*”, as séries “*Breaking Bad*” e “*The Office (UK)*” e os jogos “*The Legend of Zelda: Breath of the Wild*” (*Switch*) e “*Grand*



Figura 2.3: Cyberpunk 2077 (Playstation 4) no Metacritic.

Disponível em:<<https://www.metacritic.com/game/playstation-4/cyberpunk-2077>>. Acesso em: 26 jan. 2023

Theft Auto V” (Playstation 4) possuem estas titulações. (METACRITICS, 2021)



Figura 2.4: The Legend of Zelda: Breath of the Wild (Switch) com o título Must-Play no Metacritic.

Disponível em: <<https://www.metacritic.com/game/switch/the-legend-of-zelda-breath-of-the-wild>>. Acesso em: 13 dez. 2022

Nos últimos anos, o Metacritic se tornou um dos principais fatores para determinar quais filmes, jogos ou músicas estão em alta ou em baixa. Sendo assim, a indústria de entretenimento começou a aumentar significativamente a importância dada às avaliações e ao Metascore. Segundo Greenwood-Ericksen, Poorman e Papp (2013), produtores da área de jogos possuem uma grande preocupação com o Metascore recebido pelos projetos com que estão envolvidos. Isso ocorre devido a ampla aceitação de que o número de vendas dobra a cada 5 pontos adicionais acima da pontuação 80, ou seja, o Metascore afeta diretamente o número de cópias vendidas.

2.5 Review Bombing

Review Bombing é a expressão usada para o fenômeno em que um grande grupo de pessoas tenta realizar uma sabotagem em sites agregadores de avaliações. Este é notado quando há um alto número de submissões de críticas negativas combinadas com uma linguagem rude no comentário textual em um curto período de tempo. Sendo assim, a pontuação e classificação do item bombardeado é alterado rapidamente. (TOMASELLI; CANTONE; MAZZEO, 2021)

O termo teve origem em 2008 quando a Electronic Arts (EA) lançou o “*Spore*” com um novo sistema de gerenciamento de direitos digitais que impedia os compradores de instalar o jogo mais de três vezes. Isto foi criado como uma forma de evitar a pirataria, porém foi mal implementado pela produtora. Assim, os usuários recorreram às análises da Amazon e agregadores de avaliações com uma ampla reação negativa para o “*Spore*” e outros títulos da EA. (HANSON, 2022)

Embora as críticas desses usuários ao “*Spore*” possam parecer válidas, a cultura de *Review Bombing* tornou-se muito mais tóxica nos dias atuais com pontos de vista racistas, misóginos ou intolerantes. Como exemplo, o jogo de estratégia militar “*Total War: Rome II*” recebeu uma atualização em 2018 para permitir generais do sexo feminino e, com isso, o título sofreu diversos ataques machistas disfarçados por falta de precisão histórica. Em 2020, foi lançado “*The Last of Us Part II*” que é liderado por duas protagonistas femininas, sendo uma delas com enredo homossexual.



Figura 2.5: Avaliações do jogo Spore (PC) na Amazon.

Adaptado de: <<https://www.wired.com/2008/09/drm-opponents-a/>>. Acesso em: 26 dez. 2022

Embora seja bastante premiado, o jogo foi bombardeado de críticas homofóbicas com quase 70 mil avaliações negativas pelos usuários do Metacritic. Em 2022, a Amazon lançou a série de televisão “*The Lord of the Rings: The Rings of Power*” com elfos, anões e humanos interpretados por atores negros. Devido aos ataques racistas sofrido por conta do excesso de fanatismo dos telespectadores pela trilogia original ser esmagadoramente branca, o serviço de *streaming* decidiu suspender as avaliações dos usuários. (HANSON, 2022)



Figura 2.6: Ellie e Dina de The Last of Us Part II dançando.

Adaptado de: <<https://www.inverse.com/gaming/last-of-us-2-review-bombing-metacritic-open>>. Acesso em: 28 dez. 2022

Segundo Glennon (2020), existem seis soluções possíveis que podem evitar o

problema de *Review Bombing* no agregador de avaliações Metacritic, são eles:

- a) Possuir mais moderadores humanos: Para publicar uma avaliação no Metacritic, basta criar uma conta com um endereço de *e-mail* qualquer. Com isso, torna-se simples para o usuário participar de bombardeios de críticas negativas. É importante que existam mais moderadores humanos para que as análises sejam avaliadas e retiradas, se necessário.
- b) Incentivar os usuários: Como forma de estimular a publicação de mais resenhas, o Metacritic poderia criar uma forma de titulação para os usuários com um alto número de avaliações e comentários úteis para a comunidade. Além disso, seria viável que exista um mínimo de publicações aprovadas por um moderador para que um item seja avaliado de forma instantânea no site.
- c) Proibir avaliações antecipadas: Uma das formas de evitar o *Review Bombing* é impedir que usuários publiquem avaliações assim que um item seja criado e disponibilizado no Metacritic. Alguns jogos podem levar mais de 24 horas para terminar, então há poucas chances de que diversas pessoas avaliem logo após o lançamento. Do mesmo modo, é preciso que não seja possível fazer uma crítica antes da estreia de um filme.
- d) Confirmar que o usuário consumiu o item avaliado: Para o público que usa o Metacritic como forma de ajuda na tomada de decisão sobre como gastar seu tempo e dinheiro em entretenimento, é importante saber se o usuário que publicou a avaliação realmente consumiu o produto. Desta forma, os leitores saberiam se a opinião é de alguém que experimentou o jogo ou assistiu o filme.
- e) Tornar as escolhas do site mais transparentes: Um dos motivos para o *Review Bombing* no Metacritic é a suspeita sobre os critérios de seleção para os meios de comunicação e quais pesos são utilizados na média ponderada para o cálculo da pontuação geral. Sendo assim, o site passaria maior confiança sobre as avaliações e o Metascore.
- f) Não permitir avaliações do público: Por fim, uma forma exagerada para evitar o *Review Bombing* no Metacritic é remover completamente as avaliações de usuários. Assim, apenas as críticas de especialistas estariam disponíveis para

ser analisadas.

Capítulo 3

Proposta

Neste capítulo, será detalhado a solução proposta por este trabalho para analisar as críticas de jogos eletrônicos do agregador de avaliações Metacritic de uma forma menos sensível a ataques. Inicialmente, será apresentado a motivação pelo qual este trabalho foi desenvolvido. Em seguida, os trabalhos relacionados presentes na literatura serão apresentados. Por fim, uma descrição da solução proposta será explicada.

3.1 Motivação

Tradicionalmente, buscamos informações sobre produtos de pessoas conhecidas por meio de palavras faladas para auxiliar o nosso processo de decisão. Entretanto, sabemos que métodos como esse possuem diversos limites. Com os avanços da tecnologia, surgiu uma nova possibilidade para os consumidores realizarem consultas através de mecanismos de busca online. Assim, aumentando o conjunto de pessoas que podem colaborar nessa tarefa, uma vez que vários pontos de vista são melhores que apenas um.

Em vista disso, surgiu a ideia de criar plataformas online específicas que reunissem opiniões de diversas pessoas sobre diversos produtos. Assim, com esse objetivo, os agregadores de avaliação resumem críticas de usuários e especialistas em uma única

nota geral. Essas críticas geralmente são compostas por uma pontuação em uma escala de 0 a 10 e um conteúdo textual. Um dos sites mais conhecidos com essa ideia é o Metacritic.

Desta maneira, os agregadores conseguem ajudar consumidores a tomar uma decisão informada ao fornecer uma classificação geral e avaliações em forma de texto de outros usuários. Por um outro lado, empresas também conseguem aproveitar as avaliações positivas para ganhar credibilidade e as avaliações negativas ou neutras para aprender sobre suas falhas e necessidades (TOMASELLI; CANTONE; MAZZEO, 2021).

No entanto, esses agregadores possuem alguns problemas que são difíceis de ser resolvidos e podem comprometer a experiência do usuário. Por exemplo, a distribuição de pontuações em avaliações online possui o formato da letra J, como mostra a Figura 3.1. Isto significa que a distribuição é bimodal e distorcida em direção às notas extremas. Segundo Tomaselli, Cantone e Mazzeo (2021), isso ocorre devido ao efeito de dois vieses nos agregadores de avaliações: viés de compra e viés de subnotificação. O primeiro ocorre quando os usuários analisam todos os produtos que consomem, porém só consomem o que é bem avaliado ou popular. O segundo ocorre quando os usuários somente avaliam quando estão extremamente satisfeitos ou insatisfeitos com o item que consumiu.

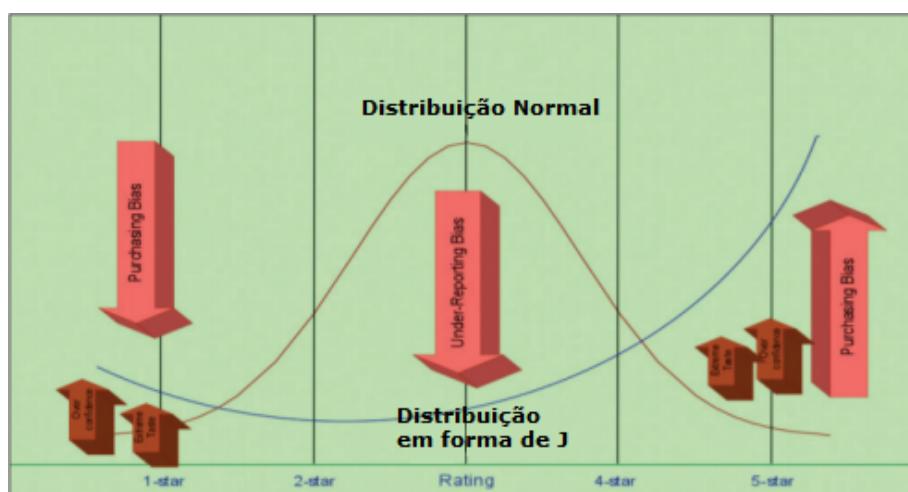


Figura 3.1: Distribuição de pontuação em avaliações online.

Adaptado de: Tomaselli, Cantone e Mazzeo (2021). Acesso em: 20 jan. 2023

Um outro problema presente nos agregadores de avaliações é o fenômeno conhecido como *Review Bombing*. Este ocorre quando um grande grupo de pessoas em um curto período de tempo realiza uma sabotagem nesses sites ao enviar várias críticas negativas combinadas com uma linguagem rude no comentário textual. Assim, a pontuação e classificação do item bombardeado é alterado rapidamente. Desta forma, o consumidor é impactado na sua decisão de compra e a forma como olha um produto. (TOMASELLI; CANTONE; MAZZEO, 2021)

O caso mais conhecido e amplo de *Review Bombing* em jogos eletrônicos ocorreu com “*The Last of Us Part II*”. Poucos dias depois do seu lançamento em junho de 2020, o jogo se tornou o mais avaliado pelos usuários do Metacritic, recebendo uma pontuação inferior a 5. Ainda que tenha sido bastante premiado e eleito o jogo do ano, a ideologia dos jogadores foi o principal motivo do bombardeio de críticas negativas.

Para ilustrar um caso de *Review Bombing* será utilizado alguns comentários extraídos do Metacritic com tradução livre sobre o jogo “*The Last of Us Part II*” (*Playstation 4*). O usuário BillyJoe33 escreveu:

“Essa história deve ser a pior que eu já vi. Nunca experimentei tantos buracos na trama e esquerdistas / politicamente corretos forçados e sem sentido em qualquer forma de entretenimento. Realmente dói dar a este jogo uma classificação tão ruim, considerando o fato de que eu realmente amei o predecessor e estava ansioso por este apenas para ficar totalmente desapontado.”.

Um outro jogador, conhecido como Dankbrownies, comentou o seguinte:

“A história é literalmente o maior f*da-se para os fãs do primeiro jogo, esta empresa estúpida tentando empurrar uma agenda para agradar o twitter, pessoas que realmente dão a este jogo um 10 podem levar uma surra de ouro”.

Já o usuário rikisjoakin disse:

“Nova Ordem Mundial. Como destruir toda uma história misturando questões políticas e videogames”.

Como o último exemplo, o usuário Latexcatdaddy avaliou o jogo da seguinte forma:

“Liberais, politicamente corretos e degeneração LGBT arruinaram uma história que já foi boa. Tudo que um dia amei foi destruído por esses pagãos.”.

Além da não garantia de que esses usuários realmente consumiram o produto, podemos perceber que todos os usuários do exemplo e muitos outros apenas criaram a conta para participar do bombardeio de críticas negativas. Assim, observa-se a necessidade de realizar uma nova análise das avaliações de jogos eletrônicos do Metacritic de uma forma menos sensível a ataques.

3.2 Trabalhos Relacionados

Em Greenwood-Ericksen, Poorman e Papp (2013), os autores apresentam que, embora produtores importantes argumentem que o Metacritic é falho em representar com precisão o valor de jogos e impacta de forma negativa a saúde da indústria, existe uma forte relação entre o Metascore e o número de cópias vendidas, independentemente do gênero ou plataforma. Ainda assim, não é possível definir se pontuações altas causam vendas altas ou ao contrário. Pode existir uma relação mais complexa envolvendo outros fatores como marketing, exposição na mídia ou viés do revisor. Apesar disso, é evidente que o Metacritic possui seu valor como uma ótima ferramenta de avaliação.

Drachen, Bauer e Veitch (2011) propõem que existe uma relação entre o Metascore e a quantidade de vezes que um jogo eletrônico é distribuído de forma ilegal. Para isso, foi utilizado uma amostra de 173 jogos lançados nos anos de 2010 e 2011. Entre esses, um total de 127 jogos e aproximadamente 12,6 milhões de acessos foram encontrados no principal canal de pirataria online, BitTorrent. Além disso, 7 dos 10 jogos mais baixados da amostra possuem uma nota superior a 75. Como resultado final, é demonstrado que as pontuações do Metacritic explicam 14% da variação de distribuições por jogo no BitTorrent.

Santos, Strohmaier e Helic (2019) apresentaram um novo estudo sobre os compor-

tamentos de especialistas e usuários do Metacritic nas análises de jogos eletrônicos. Uma das principais diferenças encontradas foi que críticas amadoras são altamente polarizadas, ou seja, as notas geralmente são zero ou dez. Por outro lado, a mídia especializada é mais equilibrada. Além disso, jogadores casuais utilizam um vocabulário emocionalmente carregado, exibindo um sentimento mais forte em relação aos especialistas.

Kasper et al. (2019) introduz um novo estudo sobre como pontuações, gêneros de jogos e textos de análise se relacionam com a utilidade da crítica para tomadas de decisão. Para esse fim, foi utilizado mais de 300 mil avaliações de jogos do Metacritic. Como resultado, foi descoberto que a nota possui um impacto positivo na quantidade de votos úteis em alguns gêneros, como *Puzzle* e *Tático*. Isso sugere que os usuários concordem com a pontuação ao invés do conteúdo textual da crítica. Para outros tipos de jogos, como *Corrida* ou *Terceira Pessoa*, a revisão textual é importante para avaliação dos usuários sobre a sua utilidade. Por fim, foi observado que o estilo de escrita e o sentimento carregado nas resenhas também é um diferencial para a definição de uma avaliação útil.

Tomaselli, Cantone e Mazzeo (2021) apresentaram um estudo de caso observational sobre o fenômeno *Review Bombing* que ocorreu nas avaliações do Metacritic no jogo “*The Last of Us Part II*” em 2020. Para isso, foi aplicado técnicas de agregação e visualização de dados para observar e obter informações sobre a evolução durante os primeiros dias de lançamento. Assim, notou-se a existência de duas tarefas principais. Uma delas é a detecção de usuários que desejam manipular o agregador de avaliações. Para detectá-los, foi observado a criação recente da conta no site e a presença de anomalias no conteúdo textual da crítica e no nome do usuário. O outro é a detecção de usuários regulares que foram influenciados pelos bombardeiros. Nesse caso, é observado a alteração do seu comportamento por conta do apoio ao bombardeio ou pela vontade de querer contrastar os comentários negativos. Como resultado, foi verificado que esses usuários possuem uma grande alteração na polaridade com o uso de sentimentos extremos no conteúdo textual da avaliação. Além disso, nota-se que a polaridade está negativamente correlacionada com a veteranidade e diversidade lexical. Por fim, podendo generalizar para qualquer jogo, os autores concluíram que

os bombardeios são realizados por motivos ideológicos.

3.3 Proposta

Conforme discutido anteriormente, os agregadores de avaliações possuem diversos problemas complexos que são difíceis de serem resolvidos e podem comprometer a experiência de um usuário. Os principais e mais discutidos são conhecidos como o fenômeno *Review Bombing* e a distribuição de pontuações distorcida em direção às notas extremas.

Em vista disso, este trabalho propõe um projeto de ciência de dados para analisar de uma nova maneira as notas e conteúdos textuais das avaliações de jogos eletrônicos do Metacritic de uma forma menos sensível a ataques. Além disso, uma nova métrica para o cálculo da pontuação geral será proposta. Como o agregador de avaliações não divulga como é calculado o Metascore e a nota de usuário, esta torna-se uma importante contribuição deste trabalho.

No primeiro instante, deve ser realizado a etapa Obter do ciclo de vida de um projeto de ciência de dados. Esta primeira etapa consiste na aquisição dos dados. Sendo assim, os dados das críticas de jogos eletrônicos como nome de usuário, data de publicação, pontuação e conteúdo textual serão extraídos de forma automática do agregador de avaliações Metacritic.

Em seguida, a etapa Manter será realizada com o objetivo de armazenar digitalmente e padronizar os dados. Desta forma, primeiramente, os dados serão armazenados em um arquivo CSV. A seguir, será feito uma limpeza nos dados para excluir as avaliações que não possuem um conteúdo textual. Além disso, para evitar ataques, deve ser realizado uma filtragem de dados em relação ao número de palavras e a riqueza de vocabulários da crítica. Esta filtragem é importante, pois nota-se que as críticas negativas feitas para sabotar um agregador de avaliações possui poucas palavras, uma linguagem rude e pobre de vocabulários.

Na terceira etapa do ciclo de vida de ciência de dados, Explorar, deve ser feito uma análise preliminar para buscar algum tipo de padrão nos dados. Assim, será

criado alguns gráficos e diagramas baseados nas pontuações atribuídas para analisar como é a distribuição de frequência das notas, a variação da média de nota durante o tempo e a quantidade de notas negativas, neutras e positivas que foram feitas durante o tempo. Portanto, serão gerados os gráficos Histograma de pontuação, Média de pontuação com desvio padrão ao longo do tempo e Avaliações por semana.

Logo após, é realizado a etapa Analisar com objetivo de obter informações mais detalhadas sobre os dados. Para isso, será utilizada a técnica de análise de sentimentos orientada a aspectos para extrair opiniões, sentimentos e subjetividades no conteúdo textual das avaliações. De maneira geral, será feito três principais tarefas: extração de aspectos, identificação de sinônimos e classificação de sentimentos.

Por fim, na última etapa do processo de ciência de dados, Interpretar, será feito a visualização dos dados para auxiliar a compreensão das informações obtidas. Dessa forma, será gerado alguns gráficos e diagramas baseados nos aspectos extraídos dos conteúdos textuais e a sua polaridade de sentimento. Assim, torna-se possível analisar os principais, mais polarizados e mais frequentes tópicos negativos e positivos sobre um jogo eletrônico. Portanto, serão gerados os gráficos e diagramas Top termos com maior polaridade de sentimento, Top termos mais frequentes, Top termos com maior valor porcentagem de frequência vezes polaridade de sentimento, Top aspectos mais frequentes e Top aspectos com maior valor porcentagem de frequência vezes polaridade de sentimento. Todos esses serão feitos de forma separada para sentimentos negativos e positivos.

Além disso, na etapa Interpretar, o jogo eletrônico irá receber uma pontuação geral que resume todas as avaliações em um único número na escala de 0 a 10. Como boa parte das críticas feitas para sabotar o agregador de avaliações possui uma pontuação negativa e um conteúdo textual com poucas informações que realmente importam devido à linguagem rude e baixa quantidade de palavras utilizadas, a métrica para o cálculo da pontuação geral proposta neste trabalho utiliza a riqueza de vocabulários como um dos parâmetros. Portanto, a pontuação geral é calculada com uma média ponderada por riqueza de vocabulário de forma que avaliações com maior diversidade lexical no seu conteúdo textual possuem um peso maior.

Capítulo 4

Experimentos

Neste capítulo, será apresentado os detalhes dos experimentos realizados no presente trabalho. Assim, será detalhado os conjuntos de dados utilizados, a metodologia aplicada e a avaliação dos resultados finais.

4.1 Base de Dados

Os conjuntos de dados utilizados nos experimentos foram extraídos do agregador de avaliações Metacritic via *Web Scraping*. Cada uma das bases de dados é composta pelo nome do revisor, data, pontuação e um conteúdo textual. Nesta seção será definido os jogos eletrônicos utilizados neste trabalho: *The Last of Us Part II* (*Playstation 4*), *Call of Duty: Modern Warfare* (*Playstation 4*), *Elden Ring* (*PC*) e *Cyberpunk 2077* (*PC*).

4.1.1 The Last of Us Part II (*Playstation 4*)

O jogo *The Last of Us Part II* da plataforma *Playstation 4* é conhecido como o maior e mais amplo caso de *Review Bombing* ocorrido em sites agregadores de avaliações. Embora tenha sido eleito o jogo do ano, o bombardeio de críticas negativas nas primeiras semanas de lançamento e a baixa pontuação dos usuários ocorreu por conta da ideologia dos jogadores.

A base de dados da crítica especializada é composta por 121 avaliações, sendo que 113 possuem pontuações positivas e 8 possuem pontuações neutras. Com isso, o título recebeu um Metascore igual a 93 e a titulação *Must-Play*, ou seja, é um jogo obrigatório de ser jogado.

Com mais de 160 mil avaliações, *The Last of Us Part II* é o jogo eletrônico mais comentado pelos utilizadores do site Metacritic. Diferentemente das análises de especialistas, o título possui uma baixa nota de usuário com uma média igual a 5.8. Em geral, a base de dados de usuários possui 84019 avaliações positivas, 8435 neutras e 68063 negativas.



Figura 4.1: The Last of Us Part II (Playstation 4) no Metacritic.

Disponível em: <<https://www.metacritic.com/game/playstation-4/the-last-of-us-part-ii>>. Acesso em: 26 jan. 2023

4.1.2 Call of Duty: Modern Warfare (Playstation 4)

O jogo eletrônico *Call of Duty: Modern Warfare* é mais um exemplo que sofreu um bombardeio de críticas negativas em sites agregadores de avaliações. Isto ocorreu por razões políticas devido a forma em que os soldados da Rússia foram representados no modo campanha. De um modo geral, os russos são retratados como “vilões”, enquanto os soldados americanos são os “heróis”.

A base de dados da crítica especializada é composta por 78 análises. Ao todo, são 67 avaliações com notas positivas e 11 com notas neutras. Dessa forma, o jogo lançado pela *Activision* em outubro de 2019 atingiu uma pontuação positiva com o Metascore igual a 80.

Com a média da pontuação dos usuários do Metacritic igual a 3.5, o segundo jogo de *Playstation 4* mais comentado de 2019 possui mais de 9 mil análises de jogadores casuais. Em geral, são 3087 críticas com notas positivas, 455 com notas neutras e 6266 com notas negativas.



Figura 4.2: Call of Duty: Modern Warfare (Playstation 4) no Metacritic.

Disponível em: <<https://www.metacritic.com/game/playstation-4/call-of-duty-modern-warfare>>. Acesso em: 26 jan. 2023

4.1.3 Elden Ring (PC)

Embora tenha sido eleito o melhor jogo do ano de 2022, *Elden Ring* possui uma avaliação ruim por parte dos usuários do Metacritic ao comparar com as pontuações de especialistas. Isto ocorreu pelo fato do título ter sofrido um bombardeio de críticas negativas no agregador de avaliação. Alguns dos jogadores reclamaram sobre problemas legítimos que jogo sofreu nas primeiras semanas de lançamento, como *bugs* e problemas de desempenho. Porém, muitos outros parecem ter avaliado

negativamente devido o jogo ser um dos mais bem avaliados de todos os tempos.

A base de dados da crítica especializada é composta por 60 avaliações. Em geral, são 57 avaliações que possuem pontuações positivas e 3 possuem pontuações neutras. Com isso, o jogo recebeu um Metascore igual a 94 e a titulação *Must-Play*.

Com mais de 5 mil avaliações, *Elden Ring* é o segundo jogo de *Playstation 4* mais comentado pelos utilizadores do Metacritic em 2022. Com a média da pontuação dos usuários igual a 6.9, o título possui 3305 críticas com notas positivas, 362 com notas neutras e 1419 com notas negativas.

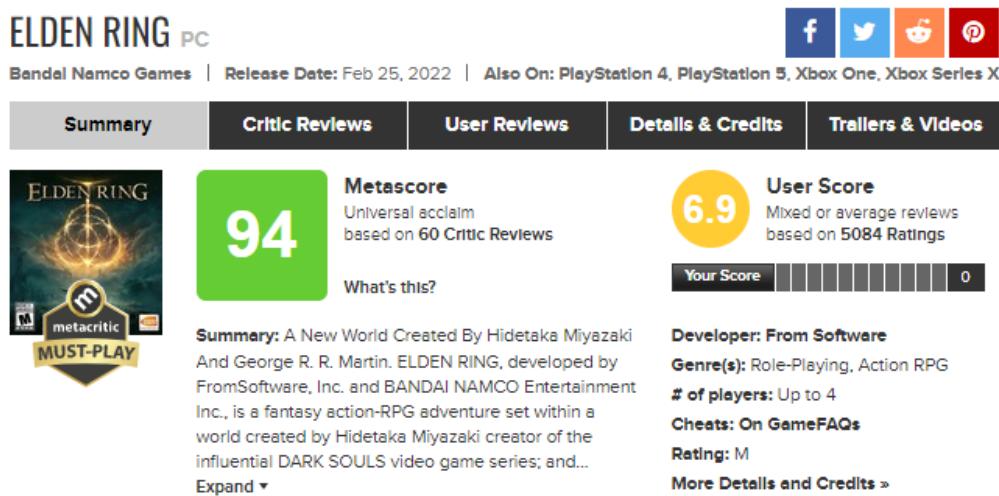


Figura 4.3: Elden Ring (PC) no Metacritic.

Disponível em: <<https://www.metacritic.com/game/pc/elden-ring>>. Acesso em: 26 jan. 2023

4.1.4 Cyberpunk 2077 (PC)

Com diversos atrasos, *Cyberpunk 2077* foi lançado em dezembro de 2020 pela *CD Projekt Red* com promessas de alcançar um novo patamar em jogos. Entretanto, as altas expectativas não conseguiram ser atendidas por conta de diversos problemas. Vários *bugs* e problemas de desempenho fizeram o jogo também ser conhecido como *Cyberbug 2077*. Além disso, as versões das plataformas *Playstation 4* e *Xbox One* precisaram ser “escondidas” pela desenvolvedora por ser impossível de jogar.

A base de dados da crítica especializada é composta por 92 avaliações. No total

são 76 avaliações que possuem pontuações positivas, 15 possuem pontuações neutras e somente 1 possui pontuação negativa. Com isso, o jogo recebeu um Metascore igual a 86.

Com a média da pontuação dos usuários do Metacritic igual a 7.0, o jogo de computador mais comentado e o segundo mais compartilhado em 2020 possui mais de 33 mil análises dos jogadores. Ao todo, são 21539 críticas com notas positivas, 2946 com notas neutras e 8599 com notas negativas.



Figura 4.4: Cyberpunk 2077 (PC) no Metacritic.

Disponível em: <<https://www.metacritic.com/game/pc/cyberpunk-2077>>. Acesso em: 27 jan. 2023

4.2 Metodologia

Nesta seção será apresentado a metodologia aplicada nos experimentos do presente trabalho. Esta foi dividida em 4 etapas principais: *Web Scraping*, Análise de Sentimento Orientada a Aspectos, Geração de Gráficos com Pontuações e Geração de Gráficos com Termos e Aspectos.

4.2.1 Web Scraping

A primeira etapa da metodologia aplicada nos experimentos consiste em utilizar a técnica *Web Scraping* para extrair de forma automática os dados do agregador de avaliações Metacritic. Inicialmente, é coletado dados como nome de usuário, data de publicação, pontuação e conteúdo textual de todas as avaliações publicadas por utilizadores do site ou especialistas. A Figura 4.5 exemplifica a extração de informações das análises da mídia especializada do jogo *God of War (PC)*.



Figura 4.5: Exemplo de Extração de Informações do jogo God of War (PC) no Metacritic.

Adaptado de: <<https://www.metacritic.com/game/pc/god-of-war/critic-reviews>>. Acesso em: 31 jan. 2023

Em seguida, para realizar uma análise de forma sensível a ataques, os dados são filtrados em relação ao seu tamanho e riqueza de vocabulário. Assim, é escolhido somente as avaliações em que o conteúdo textual da avaliação possui pelo menos 50 palavras e a riqueza lexical calculada com a métrica MATTR e tamanho de janela igual a metade de seu tamanho seja maior ou igual que 0.75. Dessa forma, é gerado a base de dados que será usada em seguida nas análises. A Figura 4.6 representa o

fluxograma da etapa de *Web Scraping*.



Figura 4.6: Etapa de Web Scraping.

4.2.2 Análise de Sentimento Orientada a Aspectos

Após a geração da base de dados, a próxima etapa da metodologia aplicada nos experimentos consiste na Análise de Sentimentos Orientada a Aspectos. Esta é responsável por extrair opiniões e determinar a polaridade do sentimento expresso nas características e atributos do conteúdo textual da avaliação.

Inicialmente, utiliza-se a técnica de Tokenização no conteúdo textual das avaliações para dividir cada texto em tokens, localizando onde começa e termina uma palavra. Em seguida, o processo de *Part-of-speech Tagging* é utilizado para atribuir uma classe gramatical a cada um dos tokens. A partir disso, é possível determinar as dependências entre cada uma das palavras do texto e, assim, definir o conjunto de aspectos e seus modificadores. Por fim, como último processo desta etapa, é aplicado técnicas de análise de sentimentos orientada a aspectos para classificar a polaridade de todos as características presentes no texto da avaliação. Desta forma, é gerado uma nova base de dados que é composta pelos seguintes dados: aspecto, modificador,

polaridade, subjetividade, data de publicação, pontuação e conteúdo textual. A Figura 4.7 representa o fluxograma desta etapa.



Figura 4.7: Etapa de Análise de Sentimentos Orientada a Aspectos.

4.2.3 Geração de Gráficos com Pontuações

A próxima etapa da metodologia aplicada nos experimentos, Geração de Gráficos com Pontuações, consiste na criação de gráficos e no cálculo de uma pontuação

geral utilizando uma nova métrica. Em um primeiro momento, de forma opcional, é realizado uma nova filtragem nos dados em relação a data de publicação das avaliações. Desta forma, torna-se possível realizar análises em um determinado período de tempo. Em seguida, é realizado algumas quantificações e cálculos para gerar uma pontuação e três gráficos. A Figura 4.8 representa o fluxograma desta etapa.

A pontuação geral é calculada com uma média ponderada por riqueza de vocabulário de forma que avaliações com maior diversidade lexical no seu conteúdo textual possuem um peso maior. O primeiro gráfico gerado é o histograma de pontuação. Este é usado para que seja possível observar a distribuição de frequências das notas. O segundo é um gráfico de linha da média de pontuação com o desvio padrão ao longo do tempo. Assim, podemos observar se a pontuação média de algum jogo aumentou ou diminuiu em um determinado dia. Por último, o terceiro gráfico representa a quantidade de avaliações feitas pelos usuários em cada semana. Estas são separadas em positivo, neutro e negativo, dependendo da pontuação atribuída.

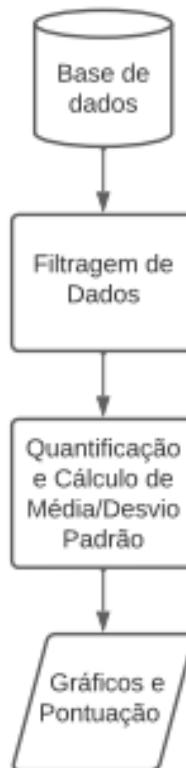


Figura 4.8: Etapa de Geração de Gráficos com Pontuações.

4.2.4 Geração de Gráficos com Termos e Aspectos

A última etapa da metodologia aplicada nos experimentos é a Geração de Gráficos com Termos e Aspectos. Esta consiste na criação de gráficos e diagramas com base nos aspectos, termos e a polaridade de sentimento. A Figura 4.9 representa o fluxograma desta etapa.

Inicialmente, como na última etapa, é realizado uma filtragem de dados em relação a data de publicação das avaliações. Em seguida, a técnica de Lematização é utilizada para determinar o lema dos aspectos com o objetivo de reunir os que possuem a mesma raiz. Além disso, é agrupado os modificadores com significados semelhantes de um mesmo aspecto. Isto é, aqueles que possuírem similaridade maior ou igual a 0.7. A partir disso, os termos e aspectos são ordenados de acordo com a polaridade de sentimento, a frequência e o valor porcentagem vezes polaridade. Assim, os seguintes gráficos e diagramas são gerados:

- a) Top termos negativos
- b) Top termos negativos mais frequentes
- c) Top termos negativos com maior valor porcentagem vezes polaridade
- d) Top aspectos negativos mais frequentes
- e) Top aspectos negativos com maior valor porcentagem vezes polaridade
- f) Top termos positivos
- g) Top termos positivos mais frequentes
- h) Top termos positivos com maior valor porcentagem vezes polaridade
- i) Top aspectos positivos mais frequentes
- j) Top aspectos positivos com maior valor porcentagem vezes polaridade

4.3 Resultados

Nesta seção será apresentado os resultados obtidos nos experimentos ao utilizar os conjuntos de dados *The Last of Us Part II (Playstation 4)*, *Call of Duty: Modern*

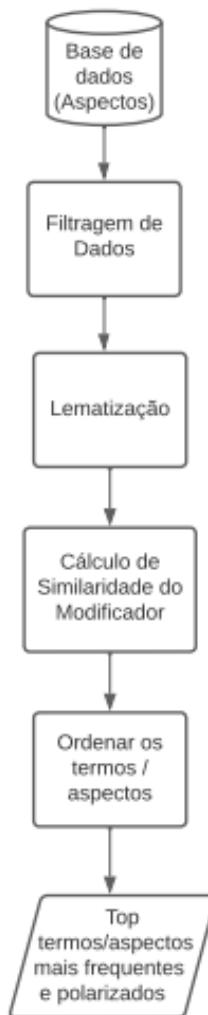


Figura 4.9: Etapa de Geração de Gráficos com Termos e Aspectos.

Warfare (Playstation 4), *Elden Ring (PC)* e *Cyberpunk 2077 (PC)*. Para isso, será feita uma análise detalhada com uma comparação entre as avaliações de usuários do Metacritic e a mídia especializada.

4.3.1 Experimento 1: The Last of Us Part II (Playstation 4)

O primeiro experimento deste trabalho tem como objetivo analisar o maior e mais amplo caso de *Review Bombing* em jogos eletrônicos ocorrido no agregador de avaliações Metacritic. O conjunto de dados do jogo *The Last of Us Part II* da plataforma *Playstation 4* possui mais de 160 mil avaliações de usuários e mais 100 avaliações de especialistas.

Em um primeiro momento, podemos observar as diferenças entre as pontuações atribuídas ao jogo pelos especialistas e usuários do Metacritic. Pela Figura 4.10, verifica-se que as notas da mídia especializada tendem a ser positivas, ou seja, entre 8 e 10. De uma forma diferente, na Figura 4.11, notamos que as pontuações dos usuários são distorcidas em direção aos extremos. Isto é, as notas tendem a ser 0 ou 10.

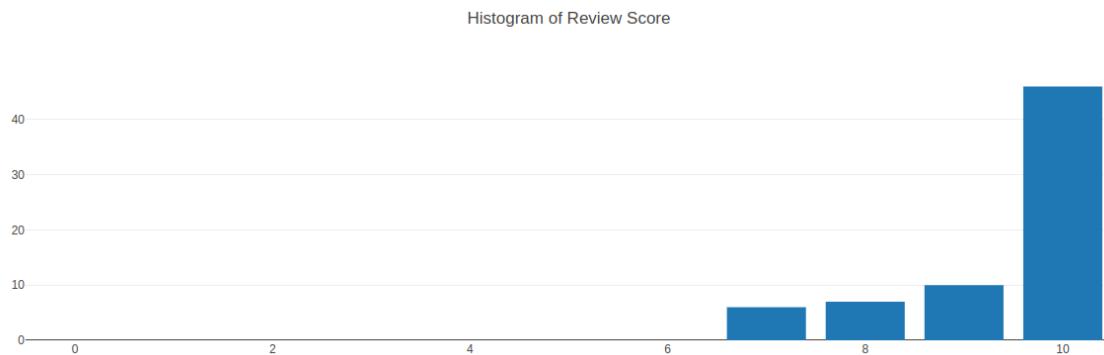


Figura 4.10: Histograma de pontuação dos especialistas.

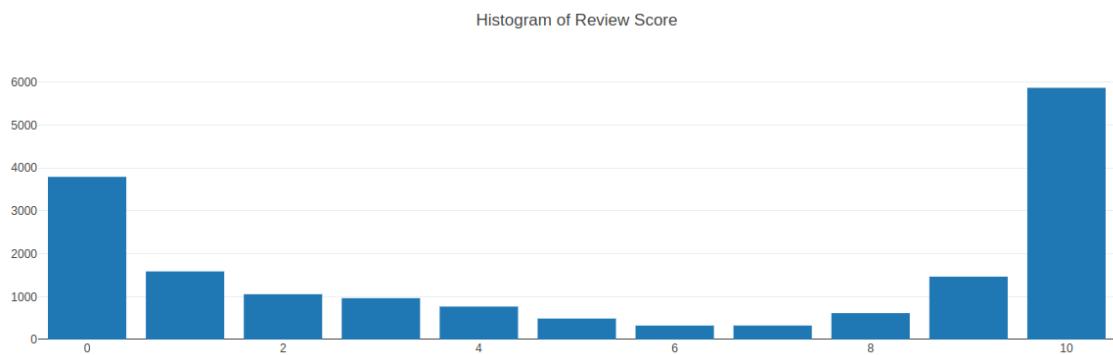


Figura 4.11: Histograma de pontuação dos usuários.

Para observar o fenômeno de *Review Bombing* que ocorreu com o jogo *The Last of Us Part II*, será utilizado um filtro de dados para selecionar somente as avaliações que foram publicadas entre 19 e 25 de julho de 2020. Na Figura 4.12, é possível perceber uma frequência maior de notas negativas. Além disso, nota-se que das quase 4 mil avaliações com pontuação 0 nos dias atuais, mais de 2 mil foram publicadas na semana de lançamento.

Em seguida, será analisado os termos mais negativos que a mídia especializada e

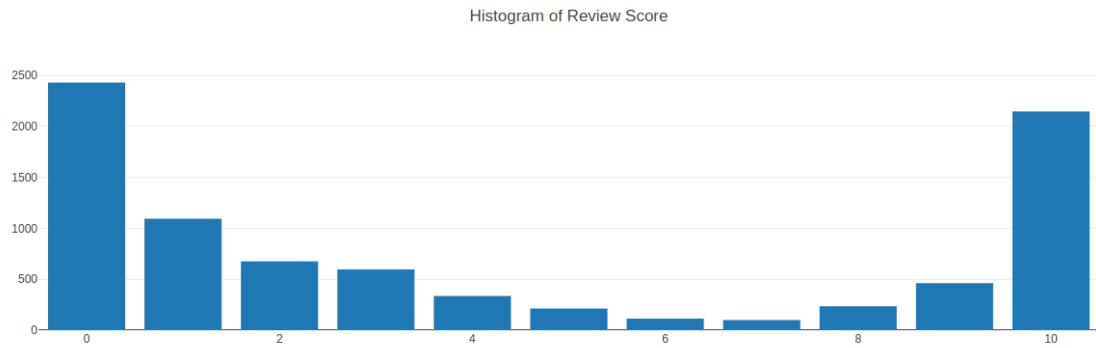


Figura 4.12: Histograma de pontuação dos usuários na semana de lançamento.

os utilizadores do site Metacritic comentaram sobre *The Last of Us Part II*. Com a Figura 4.13, notamos que os especialistas não encontraram muitos pontos negativos no jogo. Um dos tópicos comentados foram: jogabilidade sem inspiração (*uninspired gameplay*) e uma narrativa sombria (*bleak narrative*) com um comportamento inexplicável (*inexplicable behavior*) da personagem principal com ações brutais (*crude actions*) que podem não ser confortáveis para todos os jogadores.

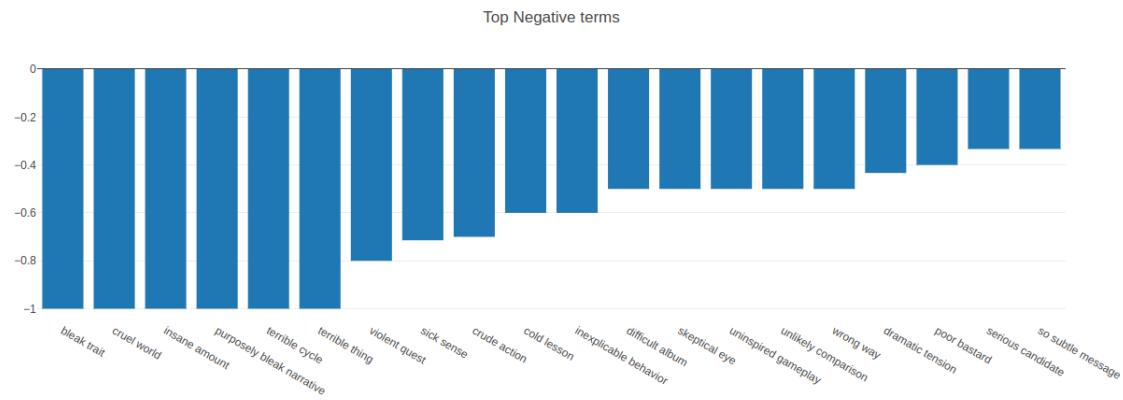


Figura 4.13: Top termos negativos dos especialistas.

De modo totalmente diferente da mídia especializada, os usuários do Metacritic apresentaram diversos pontos negativos em *The Last of Us Part II* de tal forma que não parece que o jogo tenha sido eleito um dos melhores no ano de 2020. As principais insatisfações comentadas foram sobre a história (*terrible ending*, *terrible storytelling*, *absolutely awful telling*, *absolutely terrible ending*, *absolutely terrible story*, *absolutely worst writing*), jogabilidade (*absolutely boring gameplay*), personagem principal (*terrible character*) e sua ideologia (*absolutely disgusting display of agenda*, *absolutely*

horrible insult, absolutely terrible afront).

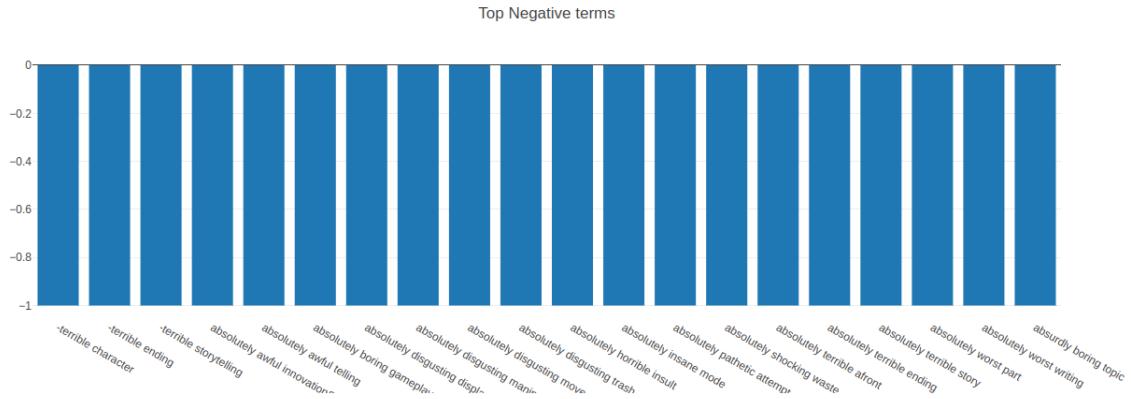


Figura 4.14: Top termos negativos dos usuários.

Isso posto, é importante destacar que boa parte dos termos negativos foram comentados em avaliações publicadas na semana de lançamento. Com as Figuras 4.15 e 4.16, notamos que isso ocorre com os seguintes tópicos: produtora Naughty Dog (*naughty dog*), história (*horrible story, awful writing, ridiculous story*) e personagem principal (*worst character*).

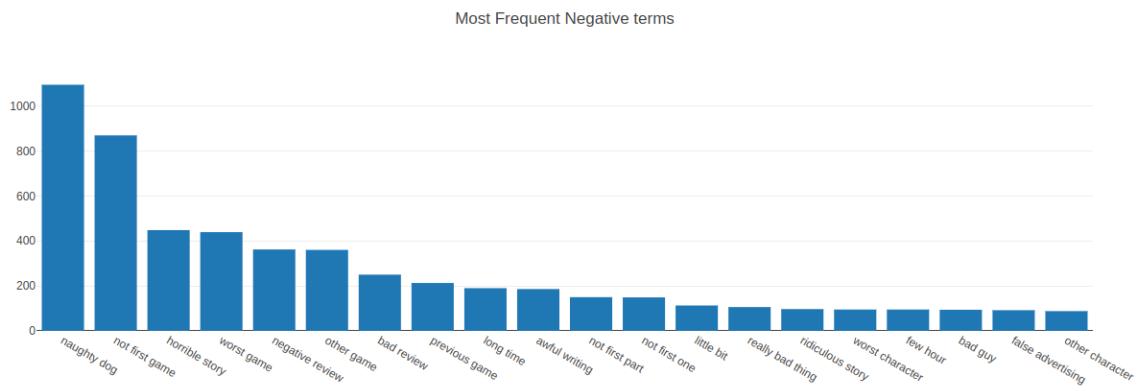


Figura 4.15: Top termos negativos mais frequentes dos usuários.

Em conclusão sobre os pontos negativos, será apresentado nas Figuras 4.17 e 4.18 os aspectos que possuem maior valor de porcentagem em que aparece no conteúdo textual das avaliações vezes a sua polaridade de sentimento. Com isso, conseguimos definir quais são os principais tópicos negativos comentados. Para a mídia especializada são: jogabilidade (*gameplay*) e história (*story*). Já para os

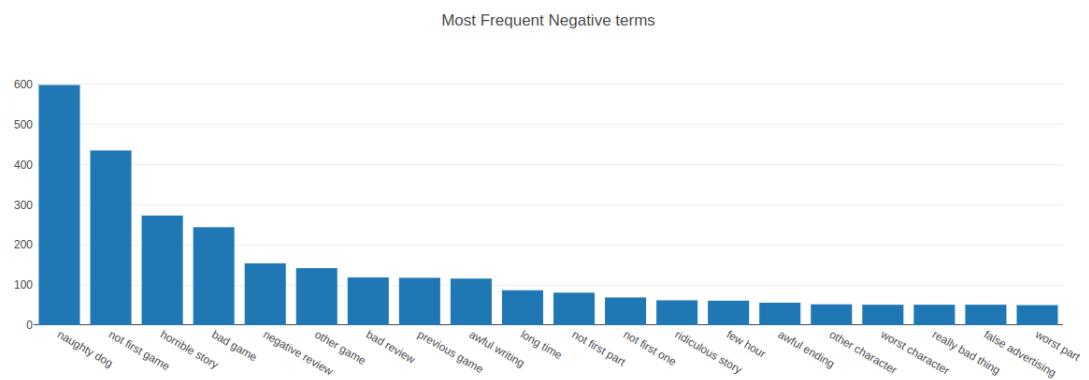


Figura 4.16: Top termos negativos mais frequentes dos usuários na semana de lançamento.

utilizadores do agregador de avaliação Metacritic são: história (*story*, *writing*), jogo (*game*) e ideologia da personagem principal (*agenda*).

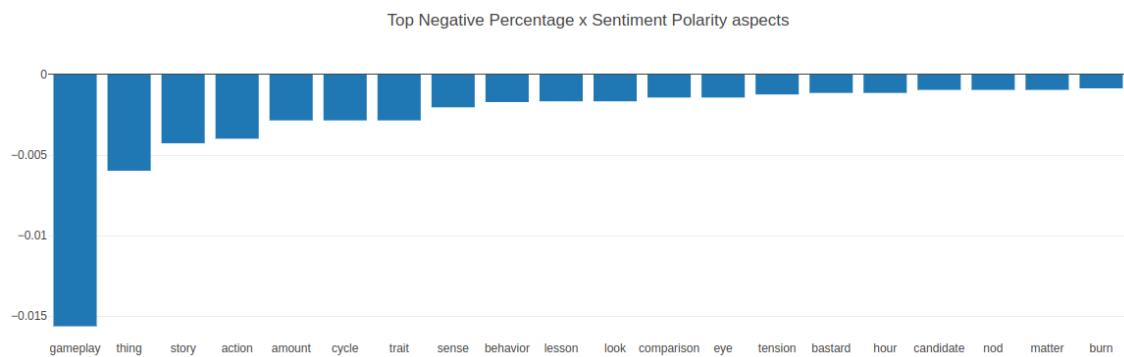


Figura 4.17: Top aspectos negativos com maior valor porcentagem vezes polaridade dos especialistas.

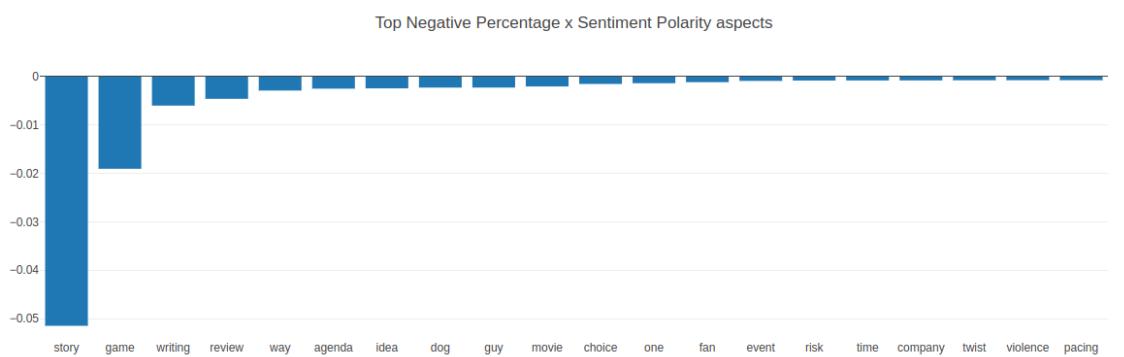


Figura 4.18: Top aspectos negativos com maior valor porcentagem vezes polaridade dos usuários.

Neste momento, será apresentado os principais tópicos positivos que especialistas e usuários do Metacritic encontraram no jogo *The Last of Us Part II* e fazer uma comparação entre eles. Na Figura 4.19, percebemos que a mídia especializada destacou os seguintes assuntos: jogo (*greatest masterpiece, excellent game, best title, best game*), história (*masterful storytelling, excellent movie, best writing*) e jogabilidade (*masterful gameplay, best gameplay*).

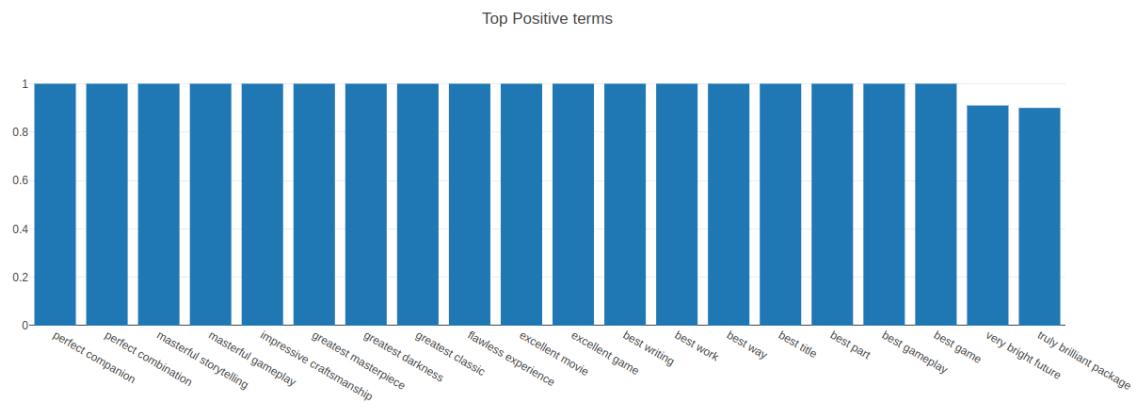


Figura 4.19: Top termos positivos dos especialistas.

Por outro lado, a Figura 4.20 evidencia que os usuários destacaram os seguintes itens: jogo (*wonderful masterpiece, wonderful game*), história (*wonderful telling, wonderful story*) e ambientação (*wonderful sound, wonderful characterization*). Desta forma, embora os pontos negativos tenham sido bem diferentes por conta do bombardeio de críticas negativas dos utilizadores do Metacritic, nota-se que os pontos positivos são semelhantes.

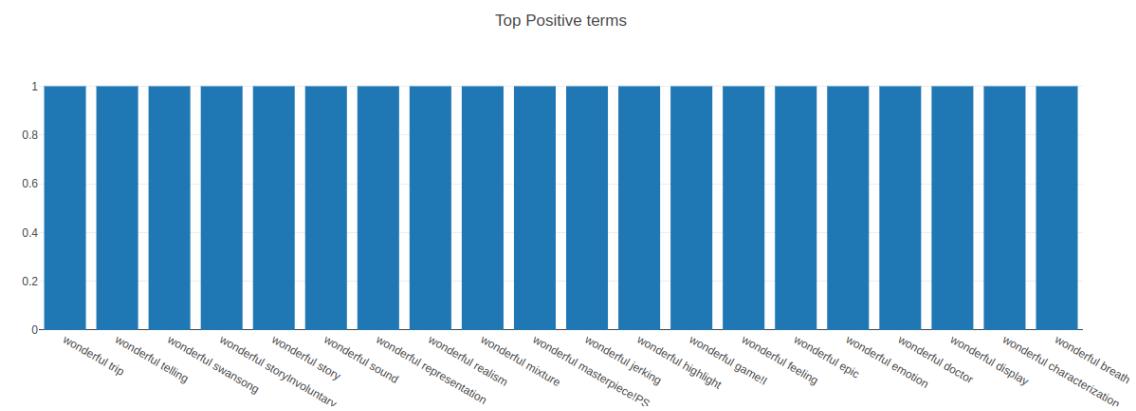


Figura 4.20: Top termos positivos dos usuários.

Para concluir sobre os pontos positivos, será apresentado nas Figuras 4.21 e 4.22 os termos que possuem o maior valor de porcentagem de frequência em avaliações vezes a sua polaridade de sentimento. Para a mídia especializada, os tópicos em destaque são: experiência (*experience*) e história (*writing, storytelling, narrative*). Já para os jogadores utilizadores do Metacritic, além de experiência (*experience*) e história (*story*), são: jogabilidade (*gameplay*), gráficos (*graphic*) e personagem principal (*character*). Assim, percebe-se mais uma vez que os pontos positivos entre os dois são comum.

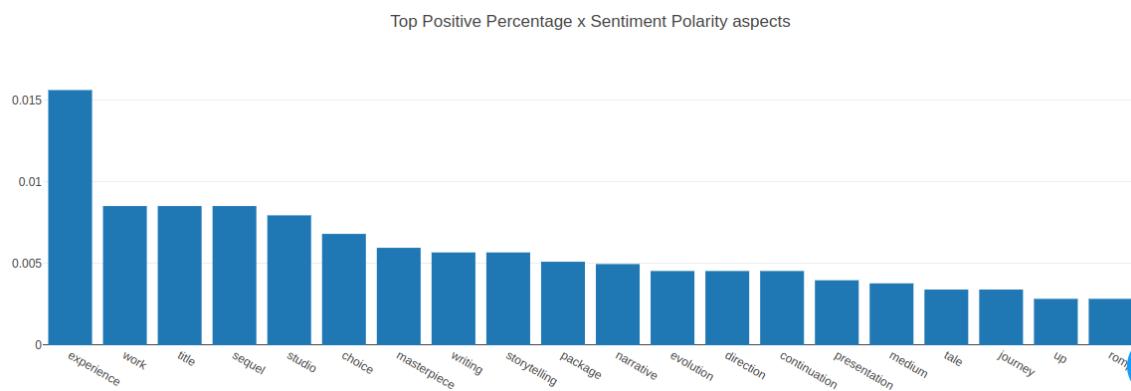


Figura 4.21: Top aspectos positivos com maior valor porcentagem vezes polaridade dos especialistas.

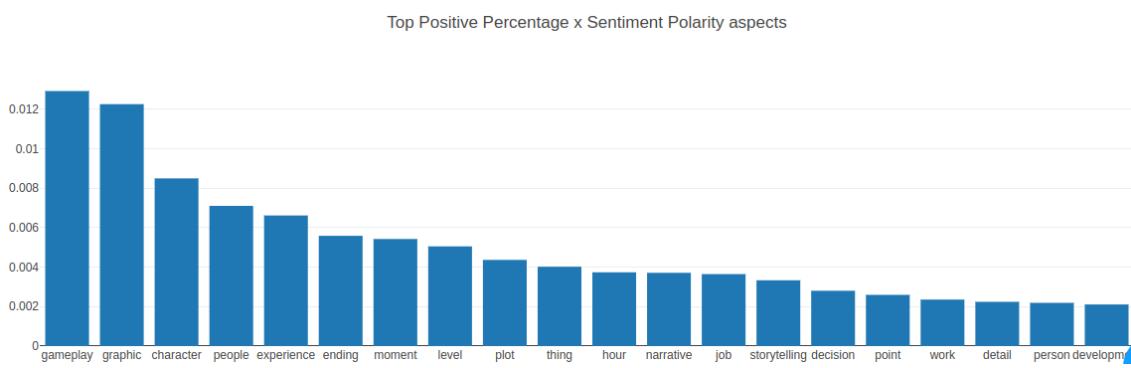


Figura 4.22: Top aspectos positivos com maior valor porcentagem vezes polaridade dos usuários.

Por fim, como última etapa da análise do Experimento 1, será apresentado a pontuação atribuída a este jogo. A média ponderada por riqueza de vocabulário obtida com as notas de especialistas foi 9.4, enquanto com notas de utilizadores do

site foi de 5.4. Como forma de comparação, o Metascore de *The Last of Us Part II* é 93 e a nota de usuário é 5.8.

4.3.2 Experimento 2: Call of Duty: Modern Warfare (Playstation 4)

O objetivo do segundo experimento deste trabalho é analisar mais um caso de *Review Bombing* em jogos eletrônicos ocorrido no agregador de avaliações Metacritic. O conjunto de dados do jogo *Call of Duty: Modern Warfare* da plataforma *Playstation 4* possui mais de 9 mil avaliações de usuários e quase 80 avaliações de especialistas.

Primeiramente, será feito uma comparação das notas atribuídas ao jogo pelos especialistas e usuários do Metacritic. Com a Figura 4.23, podemos verificar que a maioria das pontuações da mídia especializada é 8 e que nenhum deles considerou o jogo excelente com uma nota 10. De uma forma diferente, na Figura 4.24, notamos que existe uma tendência nos utilizadores do site para a pontuação 0, sendo que isso corresponde a mais do que o dobro da quantidade de notas 10.

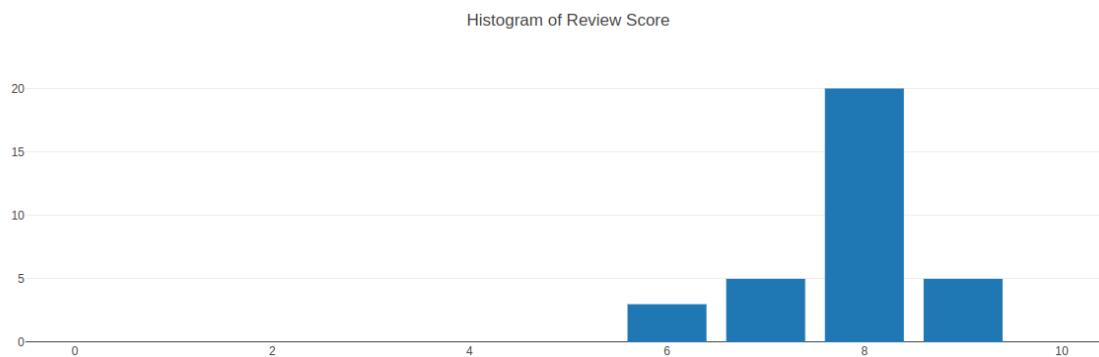


Figura 4.23: Histograma de pontuação dos especialistas.

Neste momento, será feito uma análise dos termos mais negativos que especialistas e usuários do Metacritic comentaram sobre *Call of Duty: Modern Warfare*. Com a Figura 4.25, é possível perceber os seguintes pontos negativos comentados pela mídia especializada: design de mapas ruim (*poor design*) e modo campanha (*single player*) com um péssimo enredo (*mediocre plot*) e poucas missões (*few missions*) que são muito difíceis para serem divertidas.

Da mesma forma que os especialistas, os utilizadores do agregador de avaliações

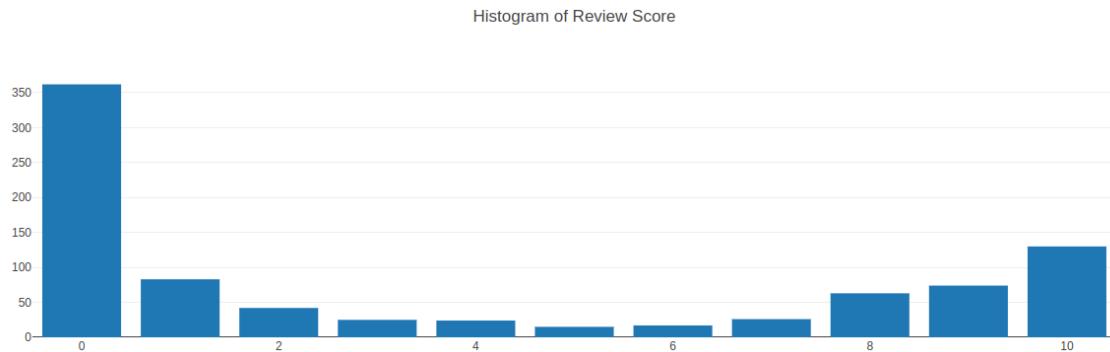


Figura 4.24: Histograma de pontuação dos usuários.

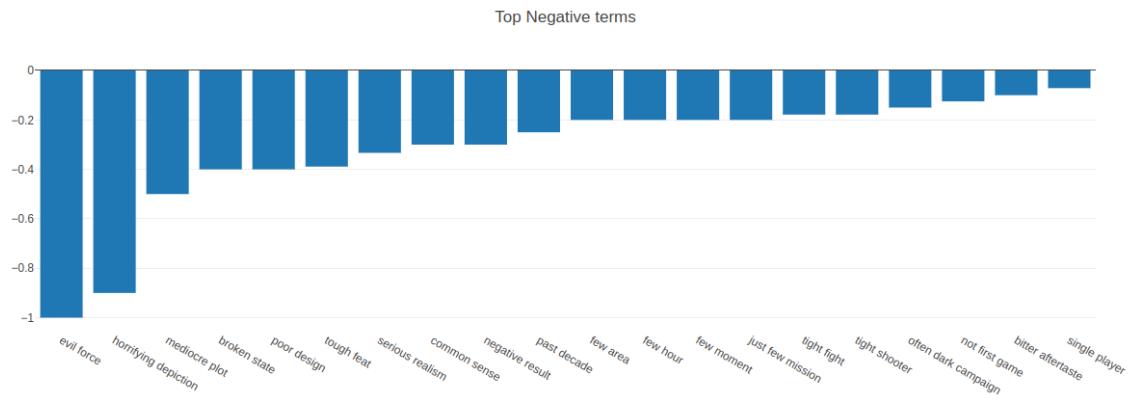


Figura 4.25: Top termos negativos dos especialistas.

Metacritic encontraram diversos problemas em *Call of Duty: Modern Warfare*. Pela Figura 4.26, podemos observar que esses ficaram extremamente insatisfeitos com o jogo (*absolutely terrible game*, *awful call of duty*, *boring game*) por conta de mapas ruins (*awful maps*), jogabilidade tediosa (*boring gameplay*), modo multijogador (*awful multiplayer*, *boring mode*) com missões não divertidas (*boring mission*) e problemas de conexão (*awful connection*, *awful server*) e, principalmente, o modo campanha que possui uma história ruim (*awful story*) que retrata os soldados russos de forma negativa (*absolutely disgusting portrayal*, *absolutely disgusting propaganda*).

Em conclusão sobre os pontos negativos do jogo *Call of Duty: Modern Warfare*, será apresentado nas Figuras 4.27 e 4.28 os termos que possuem maior valor de porcentagem que aparece nas avaliações vezes a sua polaridade de sentimento. Com isso, conseguimos concluir que existe uma similaridade entre as insatisfações de especialistas e usuários. Para a mídia especializada, os pontos negativos são: mapas

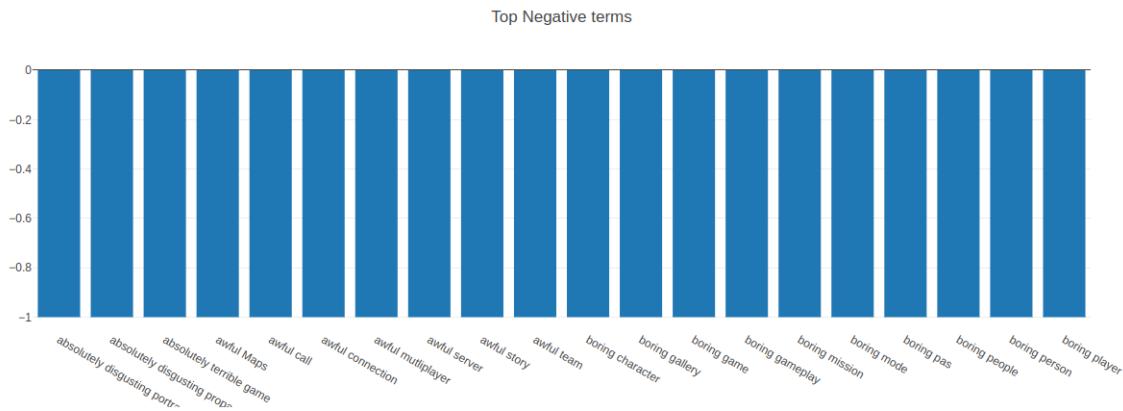


Figura 4.26: Top termos negativos dos usuários.

ruins (*poor design*) e modo campanha (*single player, often dark campaign*) com poucas missões (*just few mission*) e péssimo enredo (*mediocre plot*) pela retratação negativa dos soldados russos (*evil force, horrifying depiction*). Já para os utilizadores do site, o jogo não é bom (*bad game, awful call of duty, worst cod, absolutely terrible game*) por conta dos mapas (*worst map, horrible design*), modo multijogador (*worst multiplayer*) com problemas de conexão (*awful server*) e modo campanha com um enredo ruim (*awful story*) pela forma que os russos são apresentados (*bad guy, bad propaganda, bad thing, worst people, disgusting propaganda*).

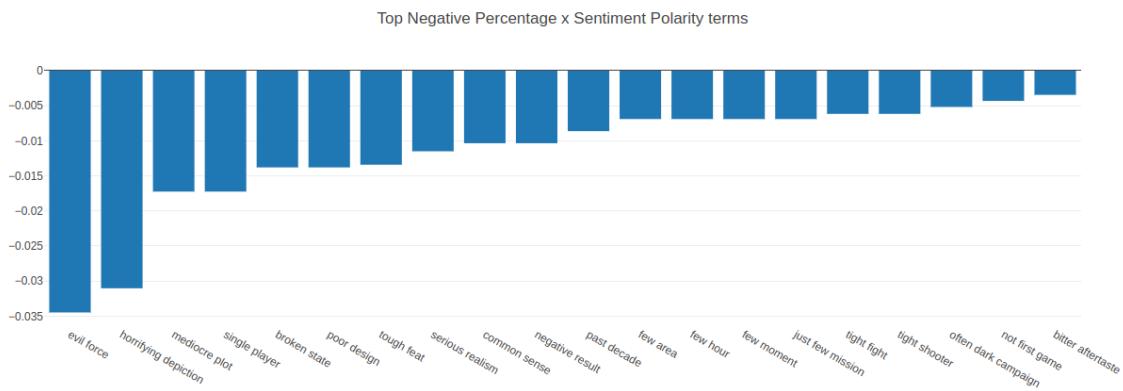


Figura 4.27: Top termos negativos com maior valor porcentagem vezes polaridade dos especialistas.

Neste momento, será apresentado os tópicos mais positivos que especialistas e usuários do Metacritic comentaram sobre *Call of Duty: Modern Warfare*. A mídia especializada o destacou como o melhor jogo eletrônico de tiro do momento

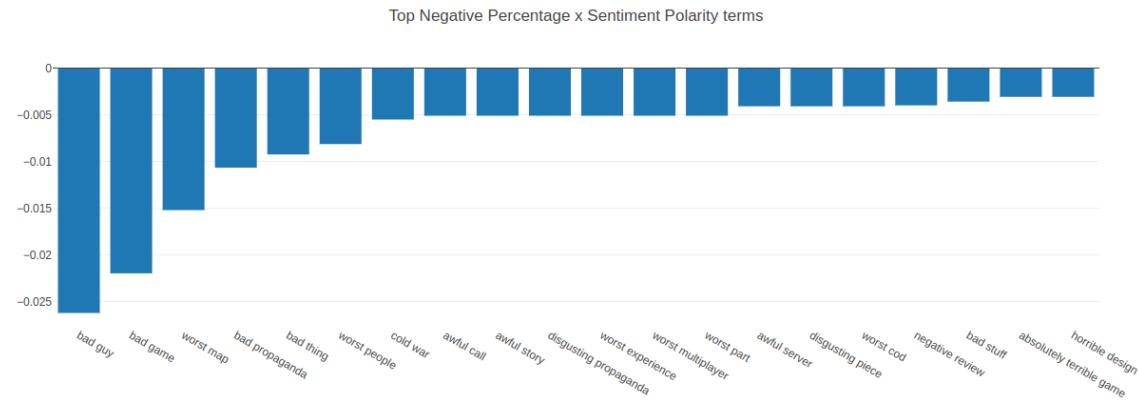


Figura 4.28: Top termos negativos com maior valor porcentagem vezes polaridade dos usuários.

(*best shooter, best game, good title*) por conta do modo campanha (*impressive campaign, great campaign, brilliant mission*) e bom visual (*great visuals, exceptional design*). Da mesma forma, alguns usuários gostaram do jogo (*perfect game, marvelous game, greatest game*) por conta dos gráficos (*best visuals*), customização (*excellent customization*), modo multijogador (*wonderful online*) e modo campanha (*greatest movie, greatest mission, greatest campaign*). Assim, nota-se que os assuntos mais agradáveis no jogo também são comum entre os dois grupos.

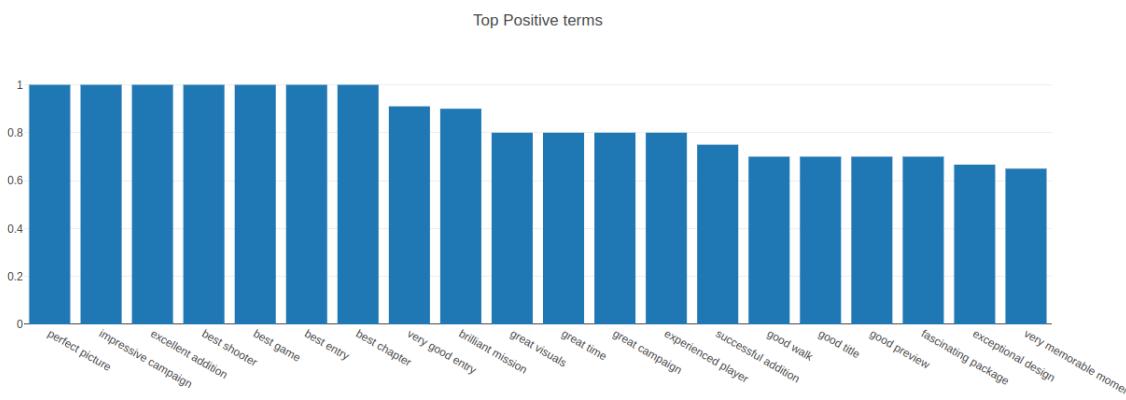


Figura 4.29: Top termos positivos dos especialistas.

Finalmente, na última etapa da análise do Experimento 2, será apresentado a pontuação atribuída a este jogo. A média ponderada por riqueza de vocabulário obtida com as notas de especialistas foi 7.8, enquanto com notas de utilizadores do site foi de 3.7. Como forma de comparação, o Metascore de *Call of Duty: Modern*

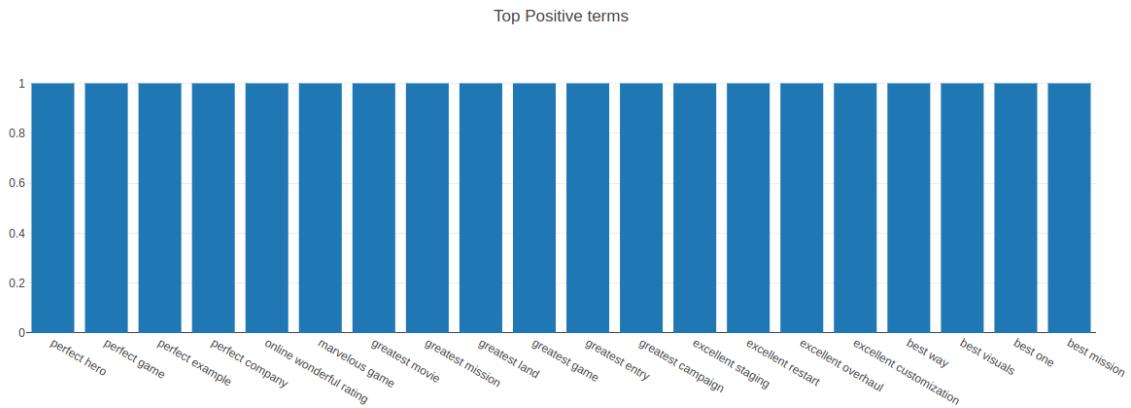


Figura 4.30: Top termos positivos dos usuários.

Warfare é 80 e a nota de usuário é 3.6.

4.3.3 Experimento 3: Elden Ring (PC)

O terceiro experimento deste trabalho tem como objetivo analisar o caso de um jogo eletrônico que foi eleito um dos melhores no ano de 2022, porém possui uma avaliação ruim de usuários do Metacritic. O conjunto de dados do jogo *Elden Ring* da plataforma *PC* possui mais de 5 mil avaliações de utilizadores do agregador de avaliação e 60 avaliações de especialistas.

Novamente, em um primeiro instante, será feito uma comparação entre as pontuações atribuídas ao jogo pelos especialistas e usuários. Pela Figura 4.31, notamos que boa parte das notas da mídia especializada são positivas, sendo que a maioria é 10. Da mesma forma, pela Figura 4.32, percebemos que existe um maior número de notas 10, mas alguns jogadores avaliaram o jogo com uma nota 0.

Neste momento, será feito uma análise e comparação das insatisfações da mídia especializada e jogadores casuais com os termos mais negativos comentados nas avaliações. Com a Figura 4.33, nota-se que os críticos comentaram que algumas falhas e limitações técnicas como modo multijogador instável (*shaky multiplayer*) e câmera frustrante (*sometimes frustrating camera*) tornaram o jogo difícil de ser recomendado na versão incial (*hard game to recommend*).

Por outro lado, percebe-se com a Figura 4.34 que os usuários encontraram muitos

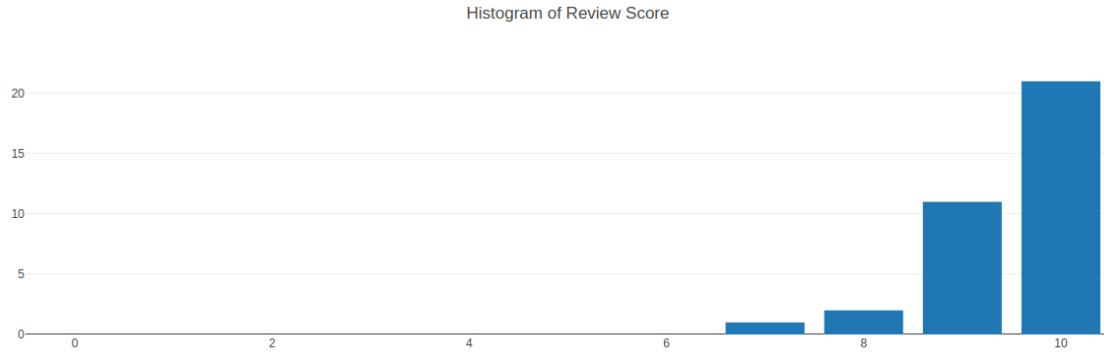


Figura 4.31: Histograma de pontuação dos especialistas.

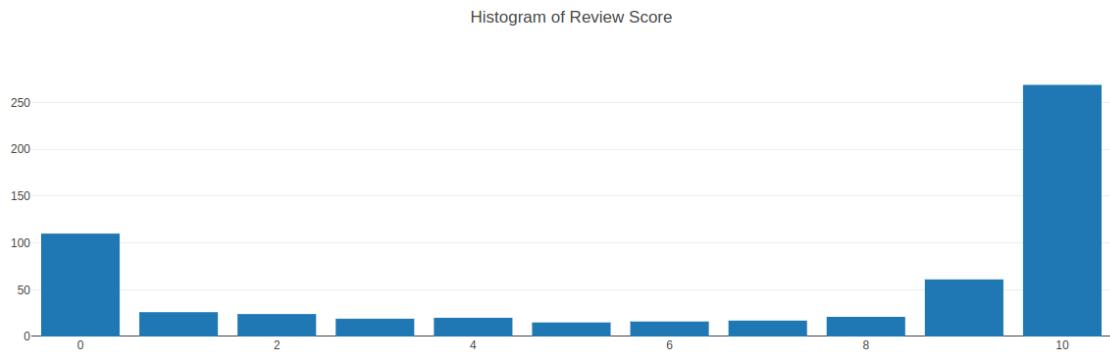


Figura 4.32: Histograma de pontuação dos usuários.

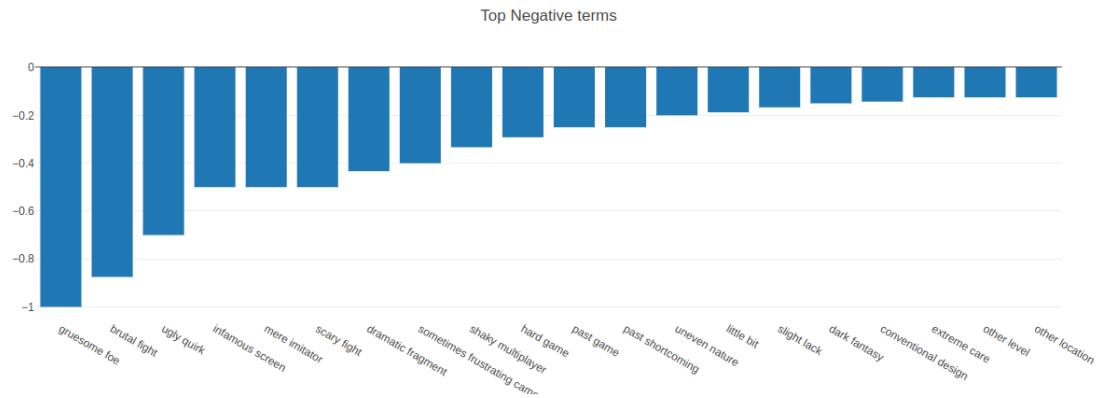


Figura 4.33: Top termos negativos dos especialistas.

mais pontos negativos que os especialistas. Muitos deles ficaram insatisfeitos com *Elden Ring* (*awful game*, *boring souls*, *boring time waster*) por conta de problemas de desempenho (*awful framedrops*, *horrible performance issues*, *horrible job*) e bugs (*awful glitch*). Além disso, foi comentado que os seguintes itens atrapalharam de

alguma forma a experiência dos jogadores (*horrible experience*): personagens (*awful character*), enredo (*boring plot, boring story*), jogabilidade (*boring gameplay, boring loop, boring repetition*), mundo (*boring map, boring world*), som (*boring music*) e menu (*horrible menu*).

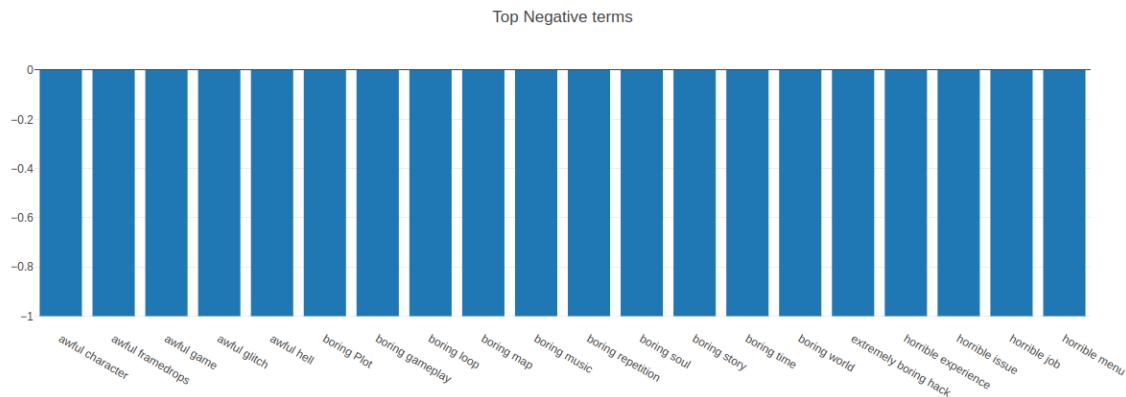


Figura 4.34: Top termos negativos dos usuários.

Para confirmar os principais pontos negativos dos usuários sobre *Elden Ring*, podemos observar o gráfico com os termos negativos que possuem o maior valor de porcentagem em que aparece nas avaliações vezes a sua polaridade, representado na Figura 4.35. Assim, verifica-se que a grande insatisfação no jogo ocorreu por conta do mau desempenho da versão para *PC* (*bad performance, terrible optimization, poor optimization*).

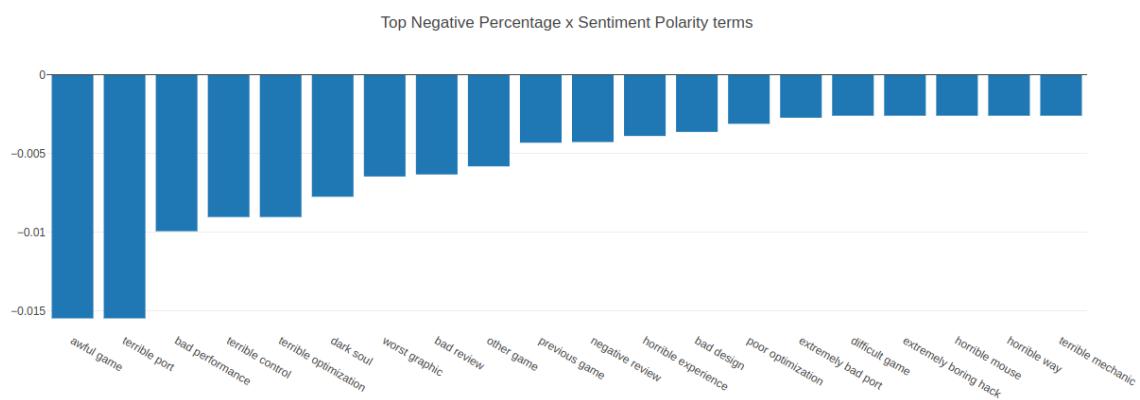


Figura 4.35: Top termos negativos com maior valor porcentagem vezes polaridade dos usuários.

Agora, será apresentado os principais pontos positivos comentados por especialis-

tas e usuários do Metacritic sobre *Elden Ring* com o apoio do gráfico que possui os termos com maior valor de porcentagem de frequência vezes polaridade de sentimento. Dessa forma, será possível entender o motivo pelo qual se tornou o melhor jogo eletrônico no ano de 2022.

Com a Figura 4.36, nota-se que os críticos consideraram o jogo como excelente (*best game, perfect game, farther best game, excellent game, incredible game*) por conta da jogabilidade (*excellent gameplay*), combate (*excellent combat*) e pelo mundo enorme que existe (*huge world*). Da mesma maneira, pela Figura 4.37, podemos perceber que os usuários definiram *Elden Ring* (*best game, amazing game, very good game, greatest game, excellent game, perfect game*) como o melhor jogo no estilo RPG e *soulslike* (*best soul, best rpg*) por conta do mundo (*amazing world, good world, beautiful landscape*), enredo (*good story*) e combate (*great combat*). Com isso, conseguimos perceber que os assuntos pontos sobre o jogo comentados por especialistas e jogadores casuais são similares.

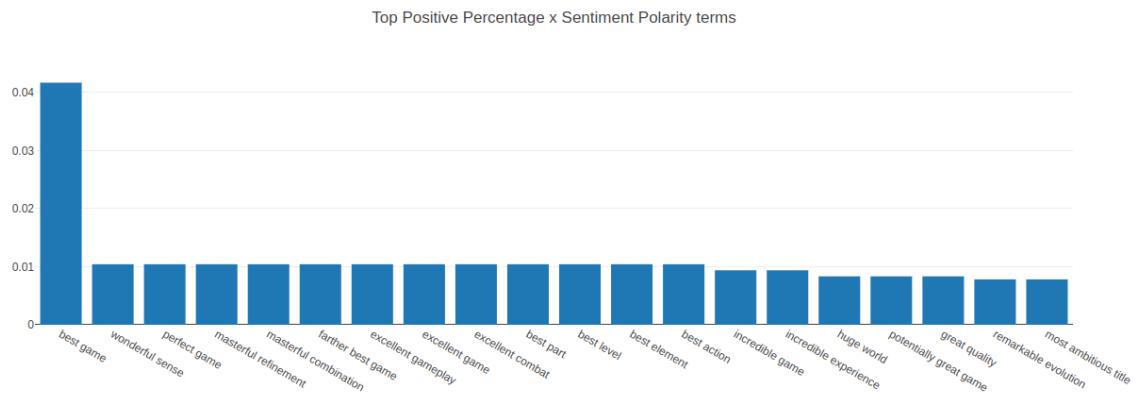


Figura 4.36: Top termos positivos com maior valor porcentagem vezes polaridade dos especialistas.

Por fim, como a última etapa da análise do Experimento 3, será apresentado a pontuação atribuída a este jogo. A média ponderada por riqueza de vocabulário obtida com as notas de especialistas foi 9.5, enquanto com notas de utilizadores do site foi de 6.5. Como forma de comparação, o Metascore de *Elden Ring* é 94 e a nota de usuário é 6.9.

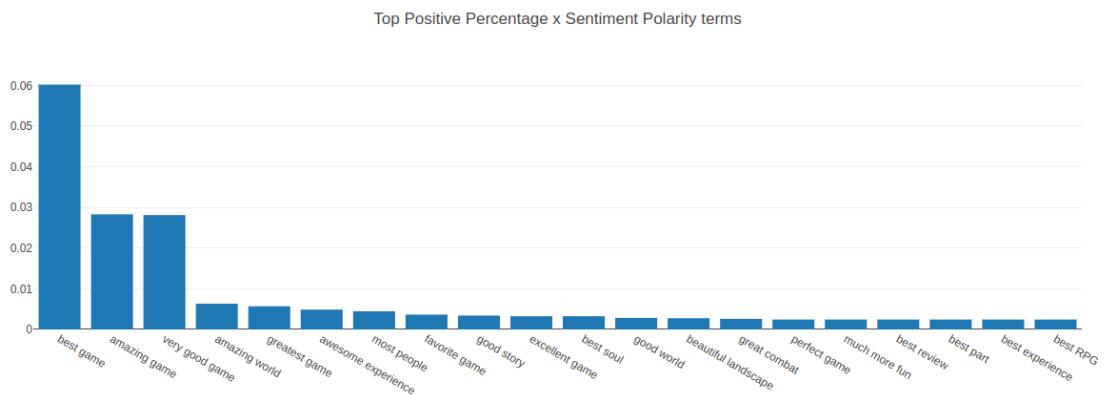


Figura 4.37: Top termos positivos com maior valor porcentagem vezes polaridade dos usuários.

4.3.4 Experimento 4: Cyberpunk 2077 (PC)

O último experimento deste trabalho tem como objetivo analisar o caso do jogo eletrônico em que as altas expectativas do público geradas por propagandas exageradas da produtora não foram atendidas. O conjunto de dados de *Cyberpunk 2077* da plataforma *PC* possui mais de 33 mil avaliações de usuários e 92 avaliações de especialistas.

Em um primeiro momento, será feito uma comparação entre as pontuações atribuídas ao jogo pelos especialistas e usuários do Metacritic. Pela Figura 4.38, verifica-se que existe uma tendência para as notas da mídia especializada serem positivas, ou seja, entre 8 e 10. Por outro lado, na Figura 4.39, notamos que as pontuações dos usuários são distorcidas em direção aos extremos, ou seja, as notas tendem a ser 0 ou 10. Ainda assim, a quantidade de nota 10 é mais que o dobro de nota 0.

Neste momento, será feito uma análise e comparação das principais decepções dos críticos e jogadores casuais com os termos negativos com maior valor porcentagem de frequência vezes polaridade de sentimento. Com a Figura 4.40, percebemos que os especialistas tiveram uma experiência decepcionante (*disappointing experience, unpleasant feeling*) por conta de alguns problemas sérios presentes no jogo (*very serious issue, unplayable version, small imperfection*). Além disso, houve uma insatisfação com a mecânica furtiva (*horrible binary stealth system*) e ecossistema

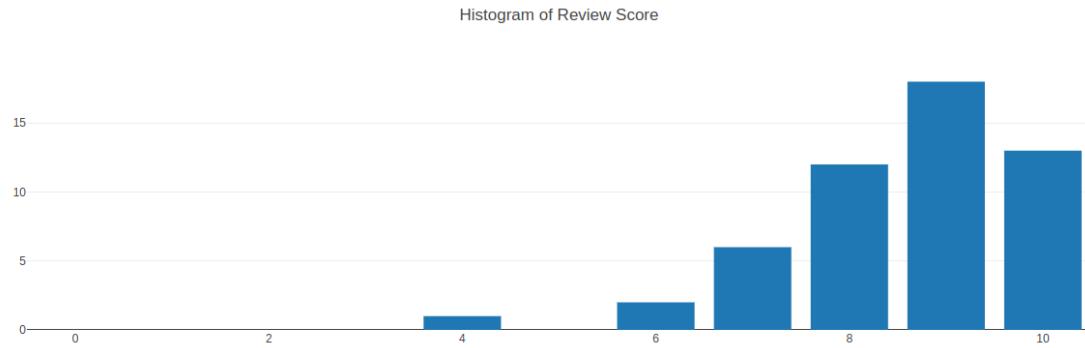


Figura 4.38: Histograma de pontuação dos especialistas.

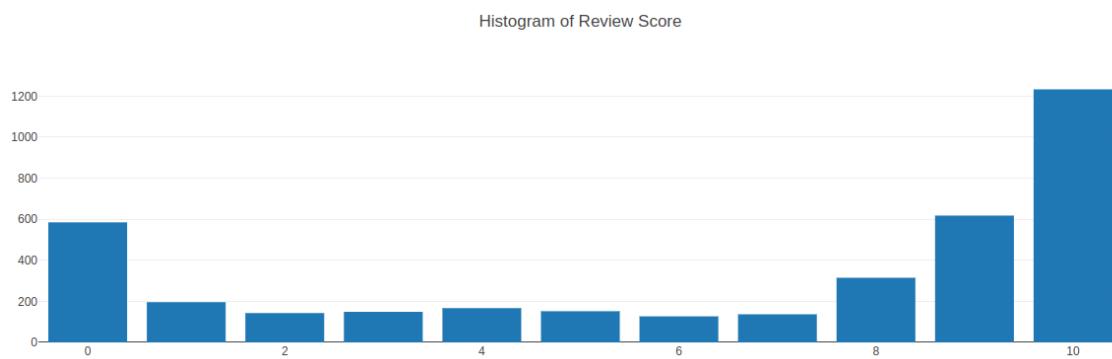


Figura 4.39: Histograma de pontuação dos usuários.

(*shallow ecosystem*).

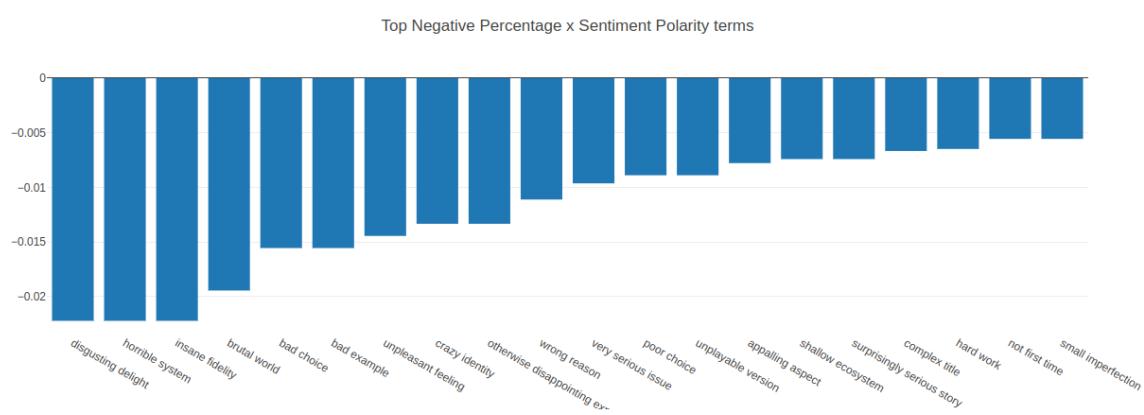


Figura 4.40: Top termos negativos com maior valor porcentagem vezes polaridade dos especialistas.

Por outro lado, pela Figura 4.41, os usuários do Metacritic encontraram diversos outros pontos negativos no jogo (*bad game*, *very bad game*, *boring game*) que

estragaram a experiência (*worst experience*). Os principais foram a quantidade de *bugs* existentes (*few bug*, *small bug*, *bad bug*) e o mau desempenho na versão de *PC* (*horrible optimization*, *terrible performance*, *poor optimization*). Além disso, houve grande decepção com o ecossistema (*worst system*), gráficos (*terrible graphic*) e condução dos carros (*terrible driving*).

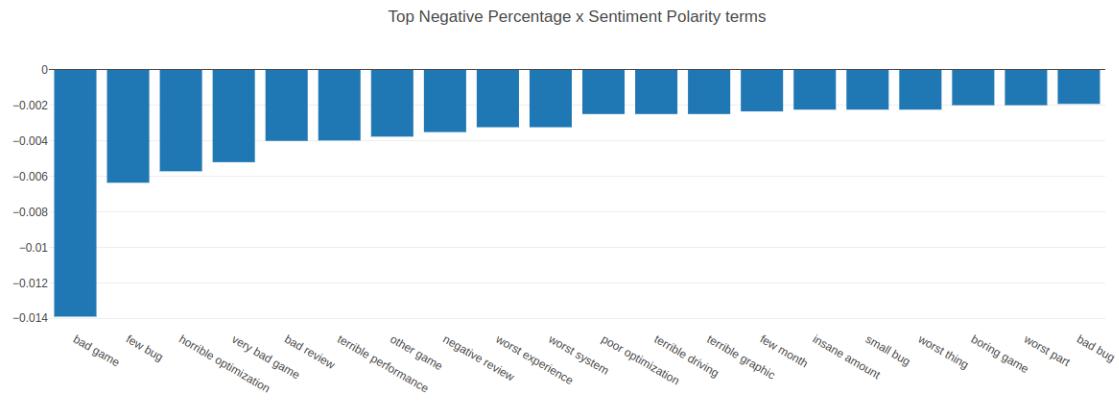


Figura 4.41: Top termos negativos com maior valor porcentagem vezes polaridade dos usuários.

Neste instante, com o apoio da Figura 4.42, será observado quais os aspectos negativos que mais se repetem nas avaliações de usuários. Com isso, conseguimos encontrar quais as principais insatisfações em *Cyberpunk 2077*. Os principais, mais uma vez, são: bugs (*bug*, *glitch*) e desempenho (*optimization*, *performance*).

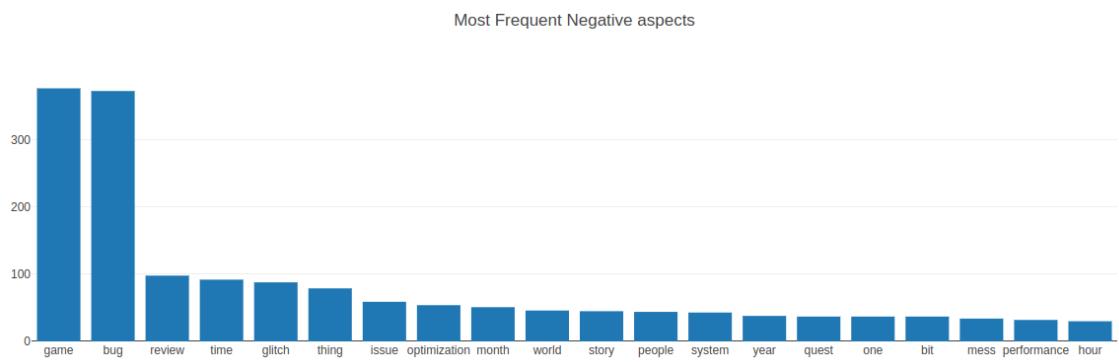


Figura 4.42: Top aspectos negativos mais frequentes dos usuários.

Agora, será apresentado os tópicos mais positivos que especialistas e usuários do Metacritic comentaram sobre o jogo eletrônico *Cyberpunk 2077*. Com a Figura 4.43, conseguimos perceber que os problemas técnicos e escolhas ruins da produtora foram

essenciais para que o jogo não tivesse um status de lendário (*truly excellent game, perfect game, legendary status, greatest game, best game*) para mídia especializada. Os principais pontos do excelente RPG (*impressive RPG, greatest RPGs*) comentados foram: gráficos (*really beautiful graphic*), enredo (*brilliant narrative*), mundo (*incredibly realized world, incredible city, brilliant world, beautiful cityscape*) e customização (*impressively flexible design*).

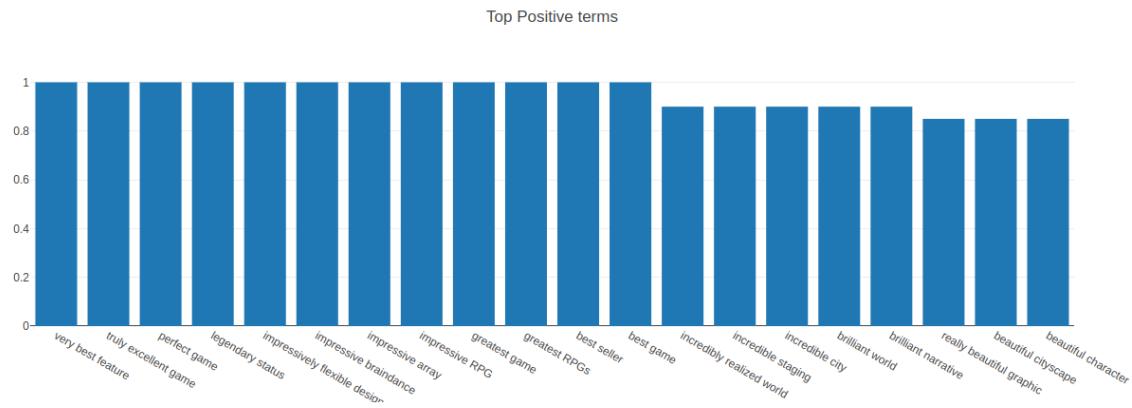


Figura 4.43: Top termos positivos dos especialistas.

Da mesma forma, alguns usuários do Metacritic avaliaram positivamente o jogo de RPG (*perfect rpg*) por conta da sua apresentação (*wonderful presentation, wonderful aesthetic*), gráficos (*wonderful excursion, visually impressive light, so awesome city*), trilha sonora (*probably best soundtrack*), história (*perfect story*) e sua profundidade (*truly impressive depth*). Assim, nota-se que os pontos positivos entre os dois são semelhantes.

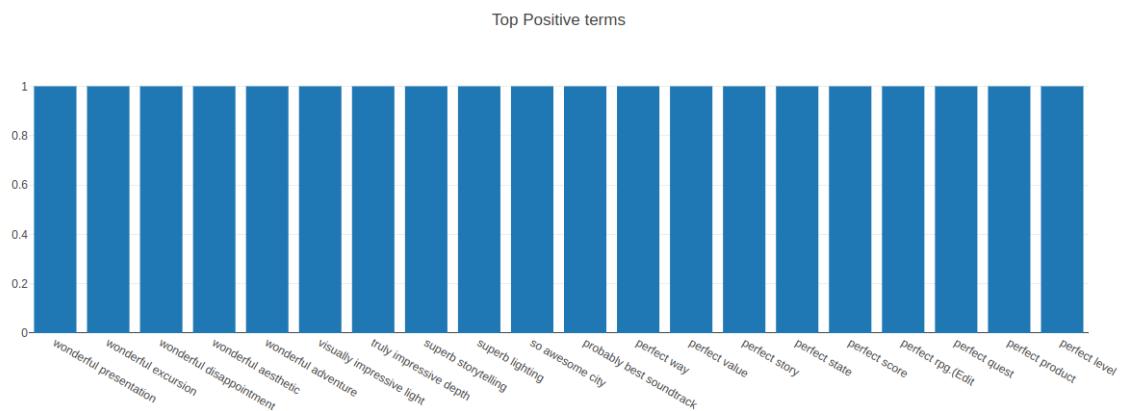


Figura 4.44: Top termos positivos dos usuários.

Por fim, na última etapa da análise do Experimento 4, será apresentado a pontuação atribuída ao jogo. A média ponderada por riqueza de vocabulário obtida com as notas de especialistas foi 8.6, enquanto com notas de utilizadores do site foi de 6.4. Como forma de comparação, o Metascore de *Cyberpunk 2077* é 86 e a nota de usuário é 7.0.

Capítulo 5

Conclusão

Neste capítulo, será apresentado um breve resumo e as considerações finais sobre o que foi feito e os resultados obtidos nos experimentos deste trabalho, junto com algumas das limitações encontradas durante o processo e possíveis extensões para trabalhos futuros.

5.1 Considerações finais

Este trabalho propôs uma nova forma de analisar os dados das avaliações do Metacritic sobre jogos eletrônicos de uma forma menos sensível a ataques e uma nova métrica para o cálculo da pontuação geral. Foram feitos diversos experimentos com bases de dados de alguns jogos que sofreram um bombardeio de críticas negativas por um motivo qualquer.

Como resultado, foi possível observar que a distribuição de pontuação dos usuários possui uma tendência em ser distorcida em direção às notas extremas. Isto significa que as pontuações dos usuários são geralmente 0 ou 10. Por outro lado, os especialistas não tem o costume de atribuir pontuações negativas. Isto pode ocorrer devido a algum fator externo não conhecido.

Além disso, nota-se que os conteúdos textuais das avaliações de usuários possuem sentimentos mais extremos e polarizados comparados aos especialistas. Em relação

aos assuntos comentados, nota-se que os pontos positivos sobre um jogo eletrônico são similares entre os dois grupos. No entanto, os usuários encontram diversos outros pontos negativos que não são comentados pela mídia especializada.

Por fim, é possível analisar que as pontuações de usuários são sempre menores que as pontuações de especialistas. Isto aconteceu por conta dos bombardeios de críticas negativas ocorridos nesses jogos no Metacritic. Além disso, podemos notar que a nova métrica para cálculo da pontuação geral proposta nesse trabalho retorna um valor próximo ao do agregador de avaliações.

5.2 Limitações e trabalhos futuros

Ainda que os resultados obtidos nos experimentos deste trabalho tenham sido satisfatórios, é possível determinar algumas melhorias a fim de aprimorar e aumentar a sua confiabilidade. Assim, esta seção tem como objetivo apresentar as limitações encontradas e possíveis progressões para um trabalho futuro.

Uma das limitações encontradas é a dificuldade em tratar e identificar sentimentos negativos nos conteúdos textuais das avaliações de especialistas e usuários. Isto é, determinar se alguma expressão teve seu significado alterado através de alguma das formas de negação explícita ou implícita. Por exemplo, o termo “fps perfeitamente estável” (*perfectly stable fps*) foi extraído no comentário de um dos usuários do Metacritic sobre o jogo *Elden Ring*. Porém, ao analisar a frase comentada, “não é possível manter 60 fps perfeitamente estáveis, mesmo em configurações baixas de 800x600” (*can't keep a perfectly stable 60 fps even on low settings 800x600*), é possível perceber o erro na identificação da negação.

Outra limitação encontrada durante o processo do trabalho é o tempo gasto para realizar um experimento, desde a etapa de *Web Scraping* até a geração dos gráficos. Como forma de exemplo, a base de dados dos usuários do jogo eletrônico *The Last of Us Part II* possui quase 800 páginas e mais de 160 mil avaliações. Dessa forma, algumas horas são utilizadas para executar todo o processo. Para trabalhos futuros, torna-se necessário melhorar o desempenho do código, seja através de otimização ou

da utilização de APIs.

Por outro lado, também podemos adicionar outros gráficos e diagramas para complementar a análise de dados das avaliações do Metacritic em um trabalho futuro. Um possível acréscimo seria utilizar alguma técnica de modelagem de tópicos para encontrar os principais assuntos que ocorrem nos conteúdos textuais das críticas. O método LDA é um exemplo que poderia ser utilizado para isso.

Uma outra possível melhoria que pode ser realizada em um trabalho futuro é no cálculo da pontuação. Neste trabalho, é feito somente uma média ponderada pela riqueza de vocabulário. Porém, uma atualização pode ser lançada para resolver alguns problemas de um jogo e, assim, algumas avaliações negativas não tenham tanto sentido a partir disso. Desta forma, podemos adicionar um novo parâmetro para o cálculo da pontuação: data de publicação.

Além dessas, uma possível melhoria que pode ser realizada em um trabalho futuro é na filtragem de dados. Neste trabalho, os dados são filtrados somente em relação ao número de palavras e a riqueza de vocabulários do conteúdo textual da avaliação. Porém, sabemos que muitos dos usuários que participam de um bombardeio de críticas negativas possuem uma conta nova e com poucas avaliações no geral. Desta maneira, podemos adicionar um filtro em relação ao número de avaliações feitas pelo usuário no agregador de avaliações.

Por fim, um outro possível direcionamento de pesquisa seria desenvolver uma análise das críticas de jogos eletrônicos obtidos em outros agregadores de avaliações, como OpenCritic. Assim, torna-se possível obter novas percepções e realizar uma comparação entre os resultados alcançados com o Metacritic e outros diferentes sites.

Referências

AWS. *What Is Data Science?* [S.l.]: Amazon Web Services, 2023. Disponível em: <<https://aws.amazon.com/what-is/data-science/>>.

BERKELEY. *What is Data Science?* [S.l.]: Berkeley School of Information, 2023. Disponível em: <<https://ischoolonline.berkeley.edu/data-science/what-is-data-science/>>.

CLEMENT, J. *Number of video gamers worldwide 2021, by region.* 2022. Disponível em: <<https://www.statista.com/statistics/293304/number-video-gamers/>>.

DASTIDAR, B. G.; BANERJEE, D.; SENGUPTA, S. An intelligent survey of personalized information retrieval using web scraper. *I.J. Education and Management Engineering*, v. 5, p. 24–31, 2016. Disponível em: <<https://www.mecs-press.org/ijeme/ijeme-v6-n5/IJEME-V6-N5-3.pdf>>.

DRACHEN, A.; BAUER, K.; VEITCH, R. W. D. Only the good... get pirated: Game piracy activity vs. metacritic score. *International Conference on Foundations of Digital Games*, v. 6, p. 292–294, Jun. 2011. Disponível em: <<https://core.ac.uk/download/pdf/17278841.pdf>>.

FERGADIOTIS, G.; WRIGHT, H. H.; WESTA, T. M. Measuring lexical diversity in narrative discourse of people with aphasia. *American journal of speech-language pathology*, 2013. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3813439/pdf/nihms517818.pdf>>.

FILHO, P. P. B. Aspect extraction in sentiment analysis for portuguese language. *Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos*, 2017. Disponível em: <https://www.teses.usp.br/teses/disponiveis/55/55134/tde-05122017-140435/publico/PedroPauloBalageFilho_revisada.pdf>.

GLENNON, J. *Metacritic has a Review Bombing problem. Here are 6 ways to fix it.* [S.l.]: Inverse, 2020. Disponível em: <<https://www.inverse.com/gaming/last-of-us-2-review-bombing-metacritic-open>>.

GREENWOOD-ERICKSEN, A.; POORMAN, S. R.; PAPP, R. On the validity of metacritic in assessing game value. *Eludamos: Journal for Computer Game Culture*, v. 7, n. 1, p. 101–127, Dec. 2013. Disponível em: <<https://septentrio.uit.no/index.php/eludamos/article/view/vol7no1-6>>.

- HANSON, S. *What is ‘Review-Bombing’? The Toxic Practice That Forced Amazon to Suspend ‘Rings of Power’ Reviews.* [S.l.]: STYLECASTER, 2022. Disponível em: <<https://stylecaster.com/what-is-review-bombing/>>.
- INDURKHYA, N.; DAMERAU, F. J. *Handbook of Natural Language Processing.* 2nd. ed. [S.l.]: Chapman & Hall/CRC, 2010.
- JURAFSKY, D.; MARTIN, J. H. *Speech and Language Processing (2nd Edition).* Prentice-Hall, Inc., 2010. Disponível em: <<https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>>.
- KASPER, P. et al. On the role of score, genre and text in helpfulness of video game reviews on metacritic. *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, p. 75–82, 2019.
- LIDDY, E. D. Natural language processing. *Encyclopedia of Library and Information Science*, p. 1–15, 2001. Disponível em: <<https://surface.syr.edu/cgi/viewcontent.cgi?article=1043&context=istpub>>.
- METACRITICS. *About Us - Metacritic.* 2021. Disponível em: <<https://www.metacritic.com/about-metacritic>>.
- MIKOLOV, T. et al. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013. Disponível em: <<https://arxiv.org/pdf/1301.3781.pdf>>.
- MIKOLOV, T.; LE, Q. V.; SUTSKEVER, I. *Exploiting Similarities among Languages for Machine Translation.* arXiv, 2013. Disponível em: <<https://arxiv.org/pdf/1309.4168v1.pdf>>.
- OUSSOUS, A. et al. Big data technologies: A survey. *Journal of King Saud University – Computer and Information Sciences*, v. 30, p. 431–448, 2018. Disponível em: <<https://fardapaper.ir/mohavaha/uploads/2018/12/Fardapaper-Big-Data-technologies-A-survey.pdf>>.
- PANG, B.; LEE, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, v. 2, n. 1-2, p. 1–135, 2008. Disponível em: <<https://www.cs.cornell.edu/home/llee/omsa/omsa.pdf>>.
- PINTO, S. C. S. Processamento de linguagem natural e extração de conhecimento. 2015. Disponível em: <<https://estudogeral.sib.uc.pt/bitstream/10316/35676/1/Processamento%20de%20Linguagem%20Natural%20e%20Extracao%20de%20Conhecimento.pdf>>.
- REINSEL, D.; GANTZ, J.; RYDNING, J. *The Digitization of the World From Edge to Core.* 2018. Disponível em: <<https://www.seagate.com/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>>.
- SANTOS, T.; STROHMAIER, F. L. and Markus; HELIC, D. What’s in a review: Discrepancies between expert and amateur reviews of video games on metacritic. *Proceedings of the ACM on Human-Computer Interaction*, v. 3, p. 1–22, Nov. 2019. Disponível em: <<https://dl.acm.org/doi/pdf/10.1145/3359242>>.

SENTIMENT Analysis Guide. Disponível em: <[https://monkeylearn.com/sentiment-analysis/#:~:text=Sentiment%20analysis%20\(or%20opinion%20mining,feedback%2C%20and%20understand%20customer%20needs.\)](https://monkeylearn.com/sentiment-analysis/#:~:text=Sentiment%20analysis%20(or%20opinion%20mining,feedback%2C%20and%20understand%20customer%20needs.))>

SIRISURIYA, S. de S. A comparative study on web scraping. *In: International Research Conference, KDU*, v. 8, 2015. Disponível em: <<http://ir.kdu.ac.lk/bitstream/handle/345/1051/com-059.pdf?sequence=1&isAllowed=y>>.

TOMASELLI, V.; CANTONE, G. G.; MAZZEO, V. The polarising effect of review bomb. *CoRR*, abs/2104.01140, 2021. Disponível em: <<https://arxiv.org/pdf/2104.01140.pdf>>.

TORRUELLA, J.; CAPSADA, R. Lexical statistics and tipological structures: A measure of lexical richness. *Procedia - Social and Behavioral Sciences*, v. 95, p. 447–454, 2013. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877042813041888>>.

Apêndice A

Mais Gráficos e Diagramas gerados no Experimento 1

Este apêndice tem como objetivo apresentar todos os gráficos e diagramas gerados no Experimento 1 em que foi usado o conjunto de dados do jogo *The Last of Us Part II* da plataforma *Playstation 4*, porém não foram utilizados na análise e avaliação dos resultados finais no Capítulo 4.

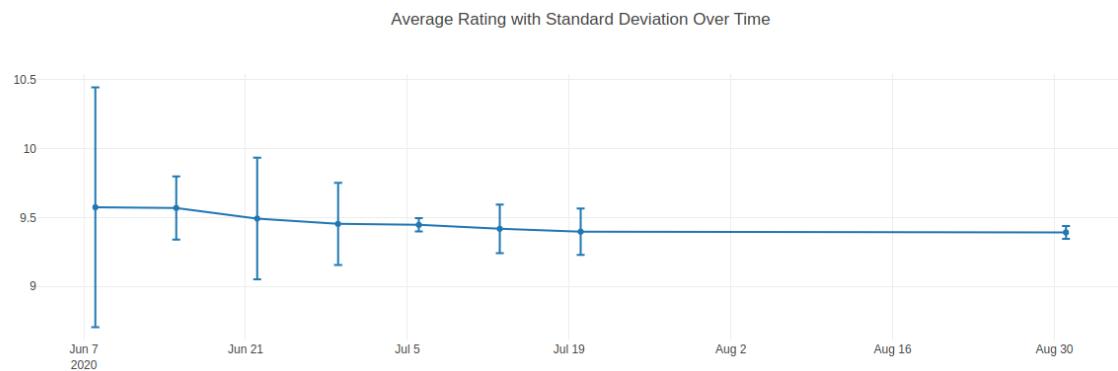


Figura A.1: Média de Pontuação com Desvio Padrão ao longo do tempo dos especialistas.

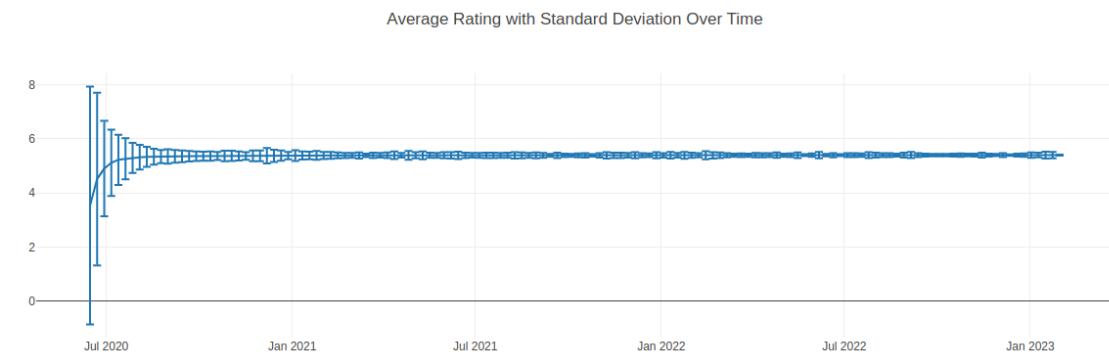


Figura A.2: Média de Pontuação com Desvio Padrão ao longo do tempo dos usuários.

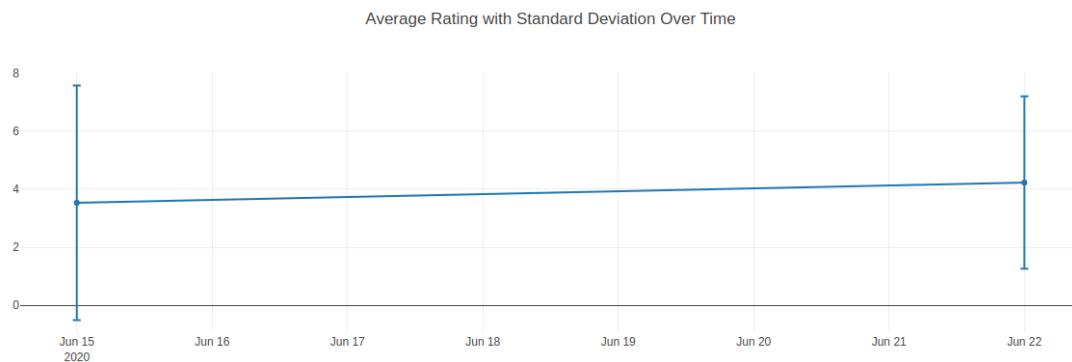


Figura A.3: Média de Pontuação com Desvio Padrão ao longo do tempo dos usuários na semana de lançamento.

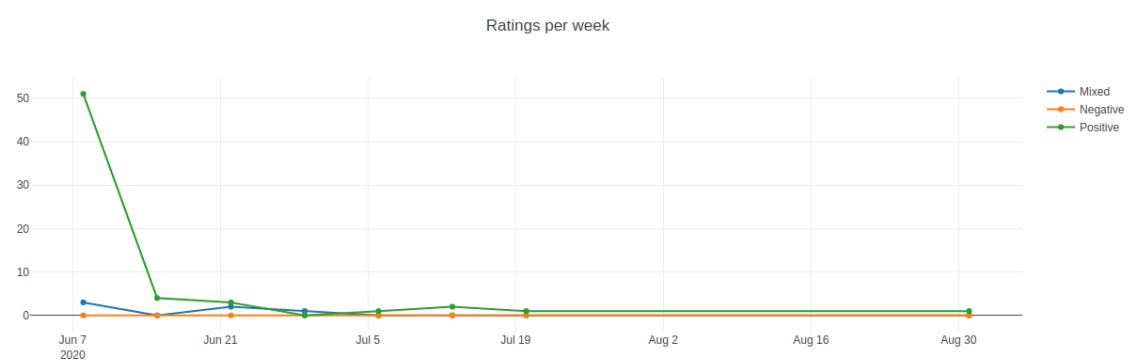


Figura A.4: Avaliações por semana dos especialistas.

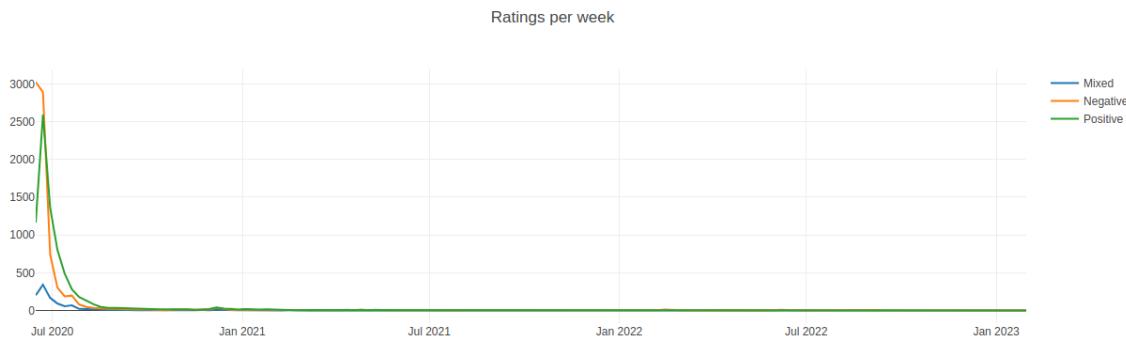


Figura A.5: Avaliações por semana dos usuários.

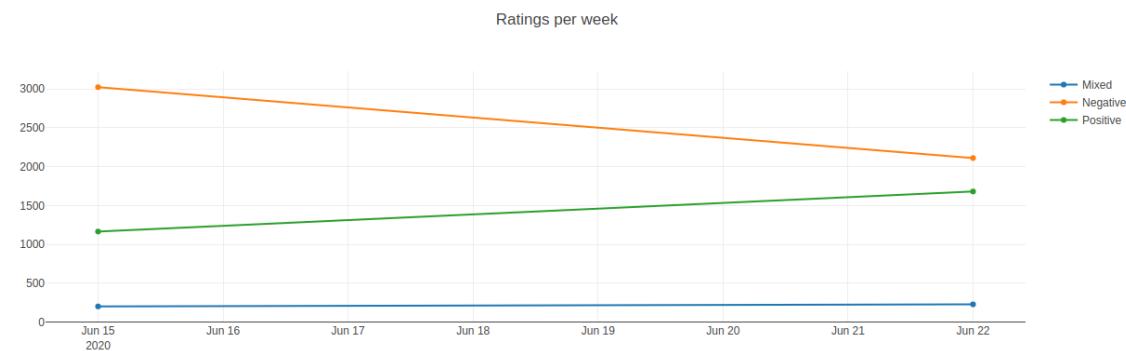


Figura A.6: Avaliações por semana dos usuários na semana de lançamento.

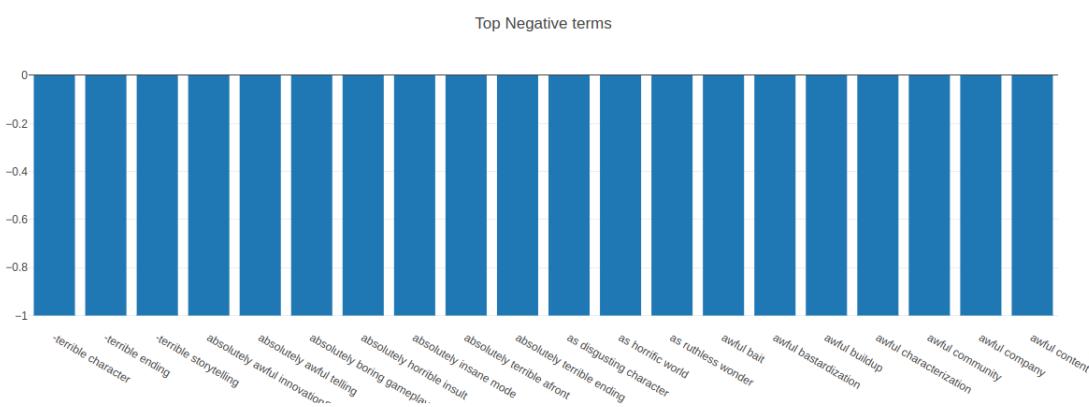


Figura A.7: Top termos negativos dos usuários na semana de lançamento.

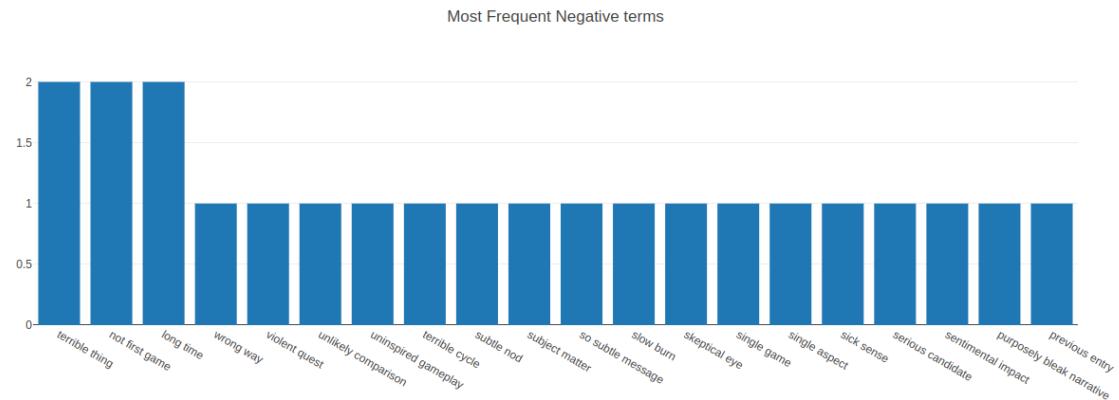


Figura A.8: Top termos negativos mais frequentes dos especialistas.

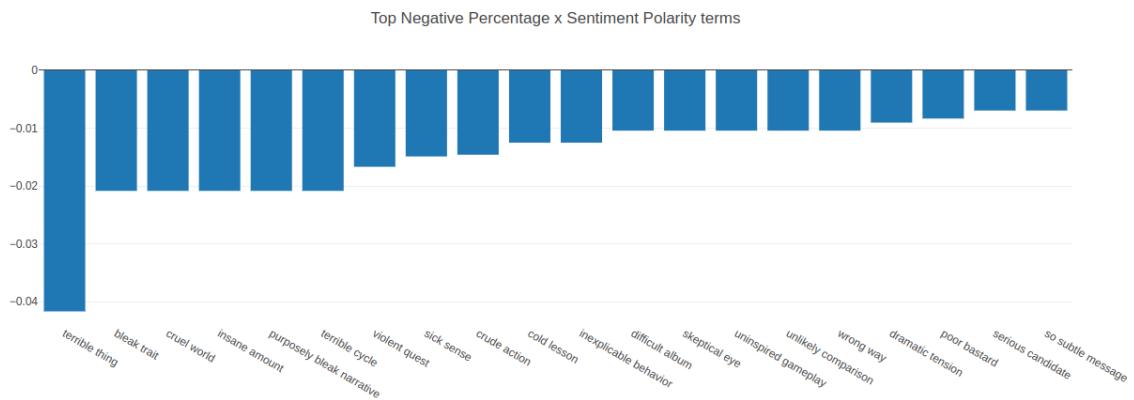


Figura A.9: Top termos negativos com maior valor porcentagem vezes polaridade dos especialistas.

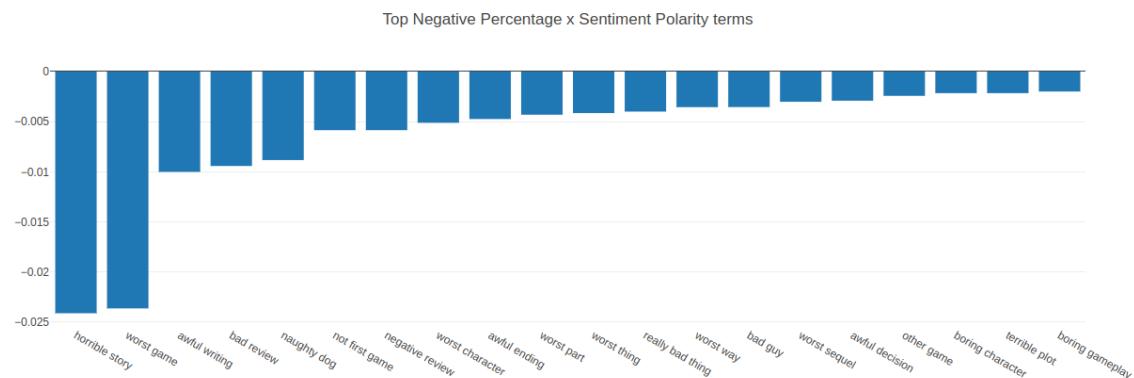


Figura A.10: Top termos negativos com maior valor porcentagem vezes polaridade dos usuários.

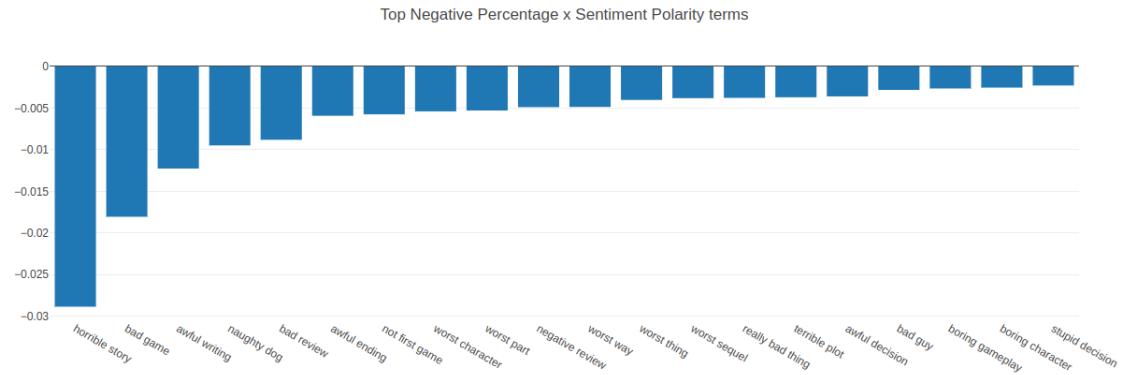


Figura A.11: Top termos negativos com maior valor porcentagem vezes polaridade dos usuários na semana de lançamento.

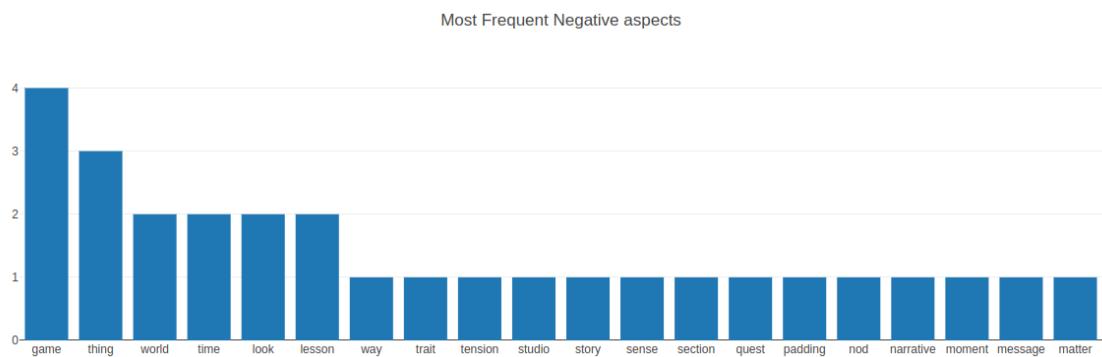


Figura A.12: Top aspectos negativos mais frequentes dos especialistas.

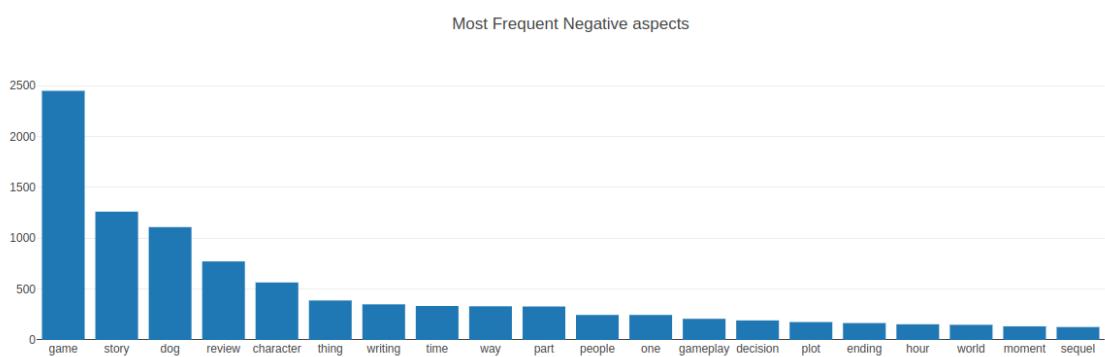


Figura A.13: Top aspectos negativos mais frequentes dos usuários.

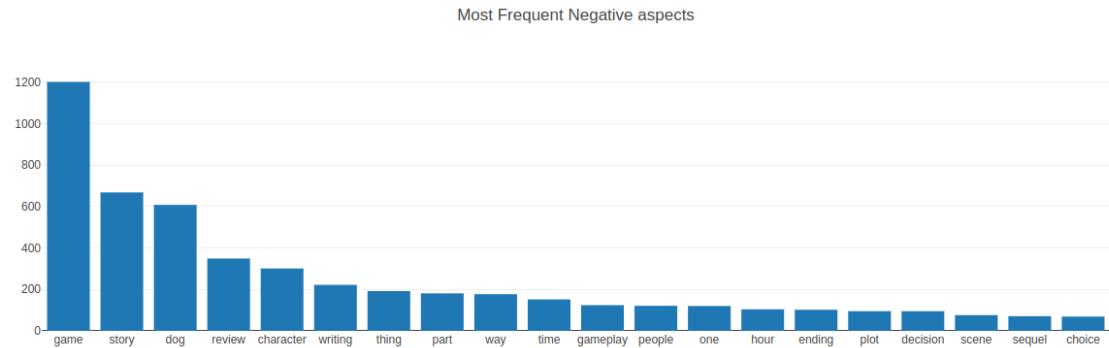


Figura A.14: Top aspectos negativos mais frequentes dos usuários na semana de lançamento.

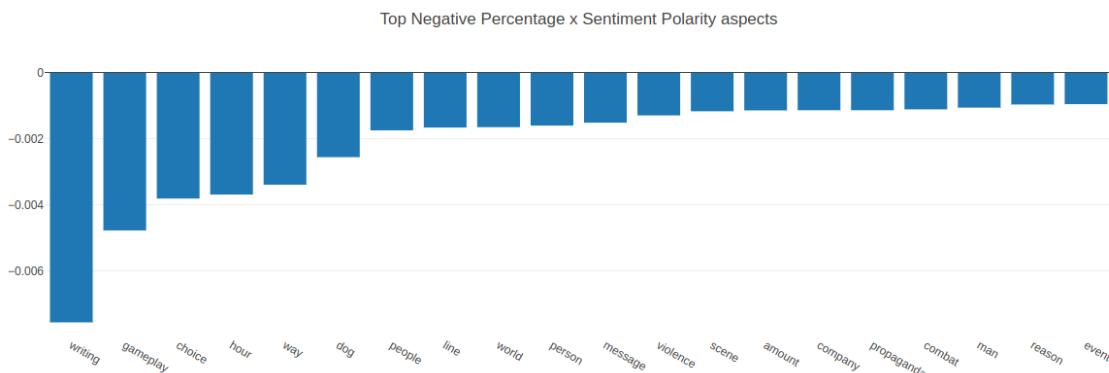


Figura A.15: Top aspectos negativos com maior valor porcentagem vezes polaridade dos usuários na semana de lançamento.

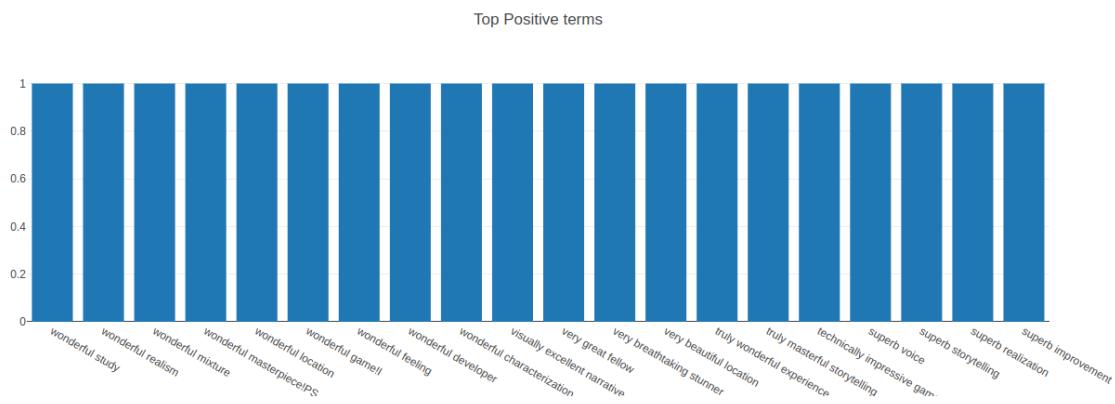


Figura A.16: Top termos positivos dos usuários na semana de lançamento.

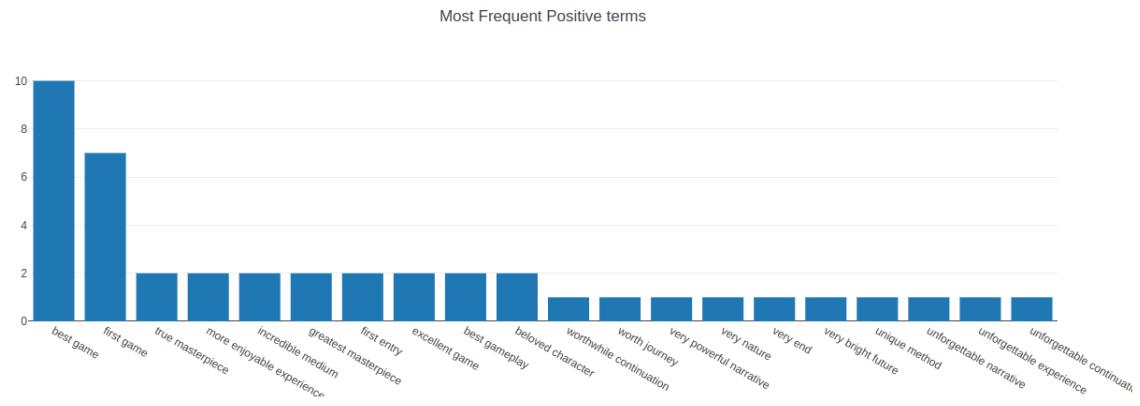


Figura A.17: Top termos positivos mais frequentes dos especialistas.

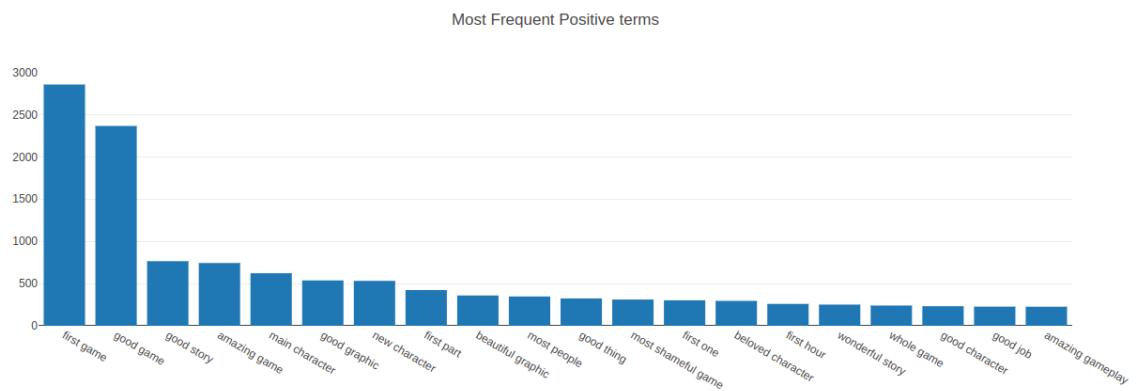


Figura A.18: Top termos positivos mais frequentes dos usuários.

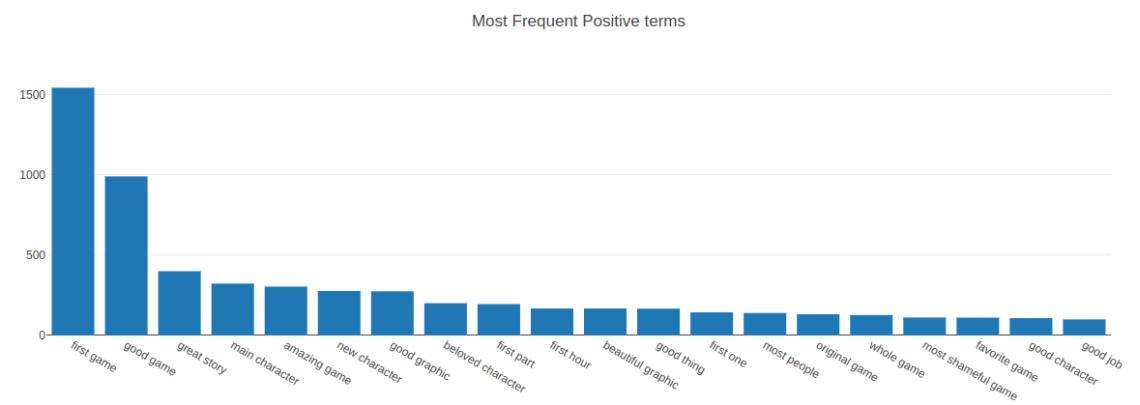


Figura A.19: Top termos positivos mais frequentes dos usuários na semana de lançamento.

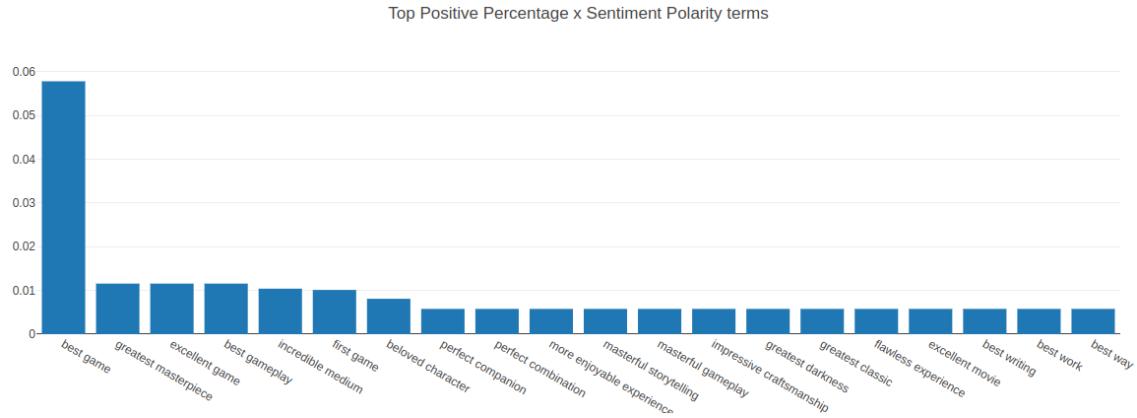


Figura A.20: Top termos positivos com maior valor porcentagem vezes polaridade dos especialistas.

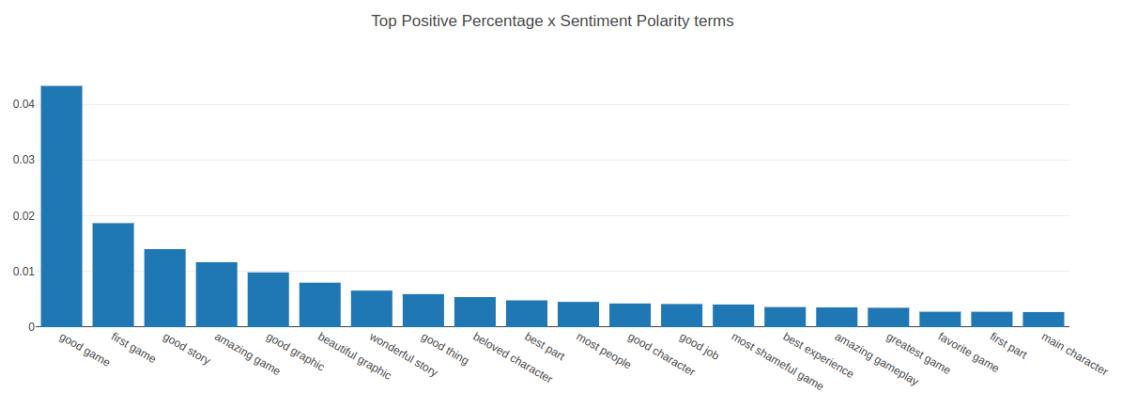


Figura A.21: Top termos positivos com maior valor porcentagem vezes polaridade dos usuários.

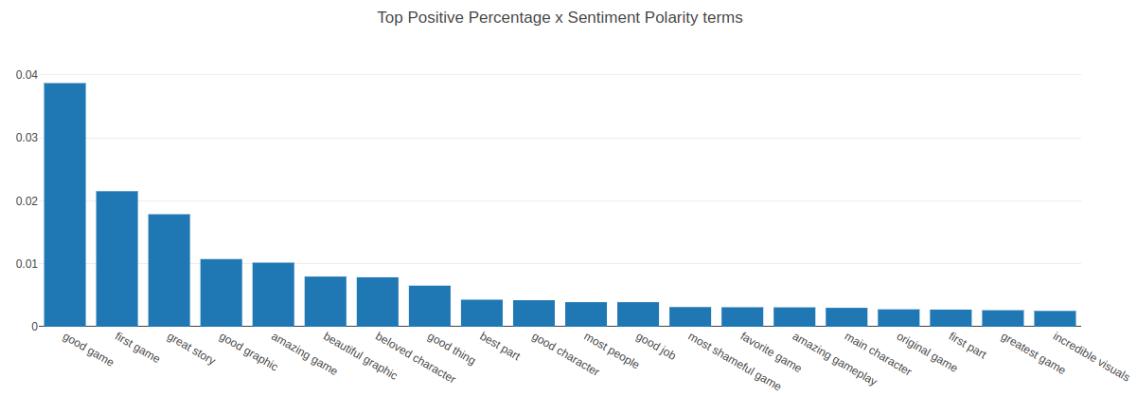


Figura A.22: Top termos positivos com maior valor porcentagem vezes polaridade dos usuários na semana de lançamento.

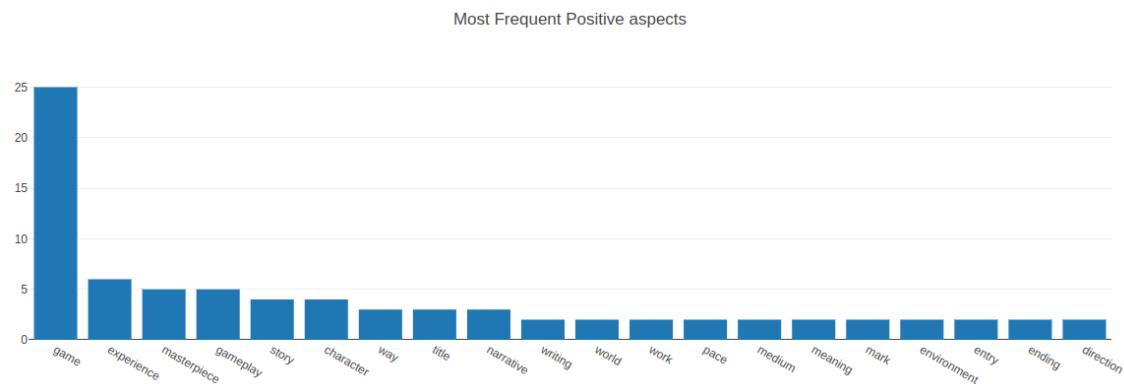


Figura A.23: Top aspectos positivos mais frequentes dos especialistas.

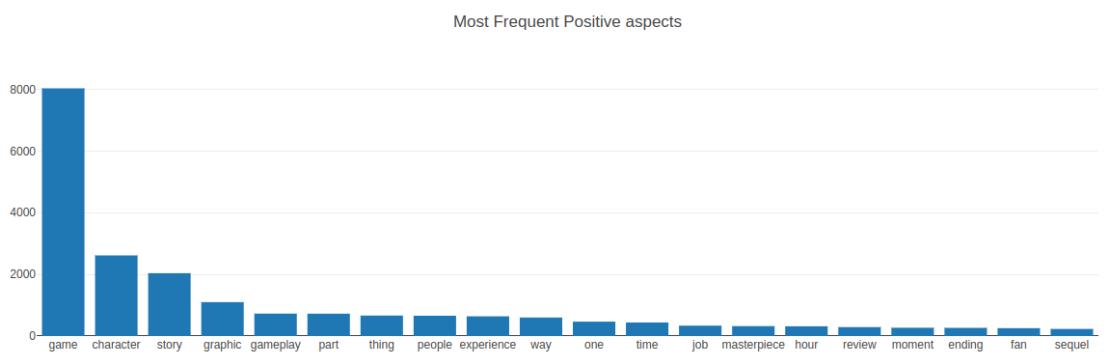


Figura A.24: Top aspectos positivos mais frequentes dos usuários.

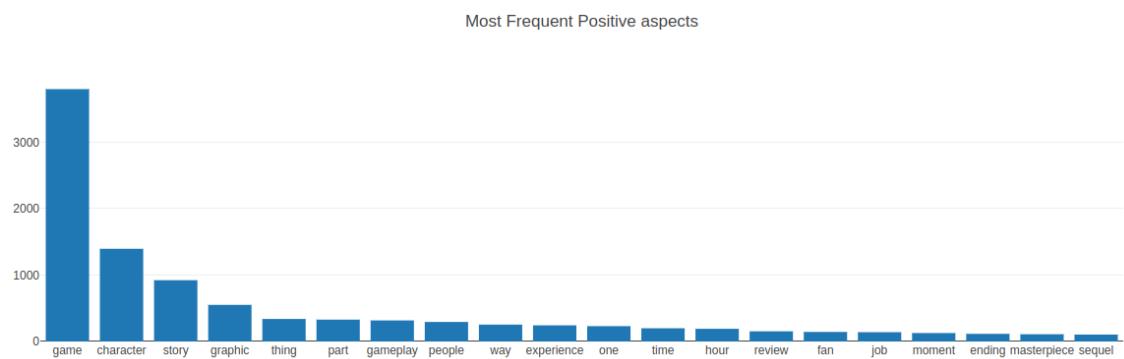


Figura A.25: Top aspectos positivos mais frequentes dos usuários na semana de lançamento.

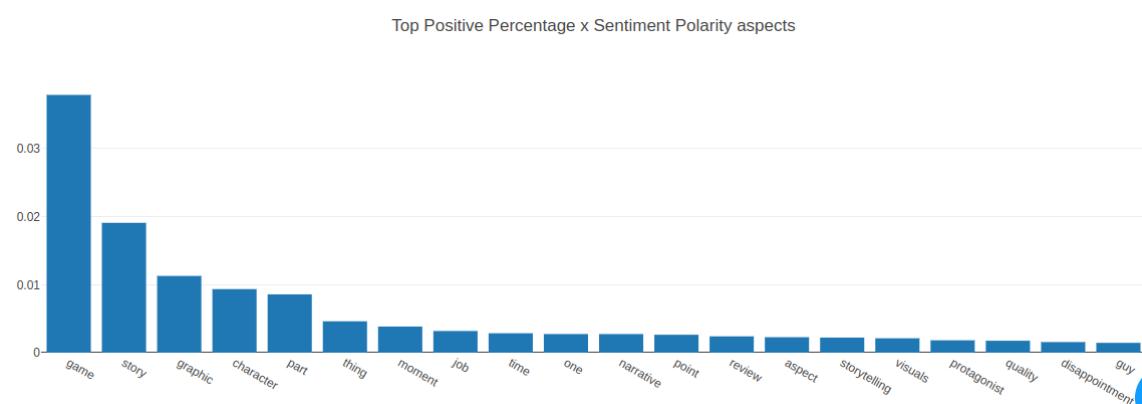


Figura A.26: Top aspectos positivos com maior valor porcentagem vezes polaridade dos usuários na semana de lançamento.

Apêndice B

Mais Gráficos e Diagramas gerados no Experimento 2

Este apêndice tem como objetivo apresentar todos os gráficos e diagramas gerados no Experimento 2 em que foi usado o conjunto de dados do jogo *Call of Duty: Modern Warfare* da plataforma *Playstation 4*, porém não foram utilizados na análise e avaliação dos resultados finais no Capítulo 4.

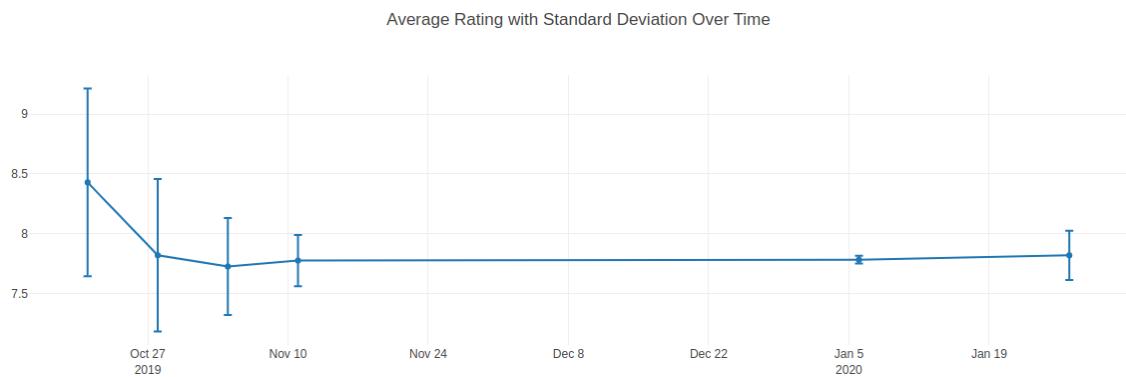


Figura B.1: Média de Pontuação com Desvio Padrão ao longo do tempo dos especialistas.

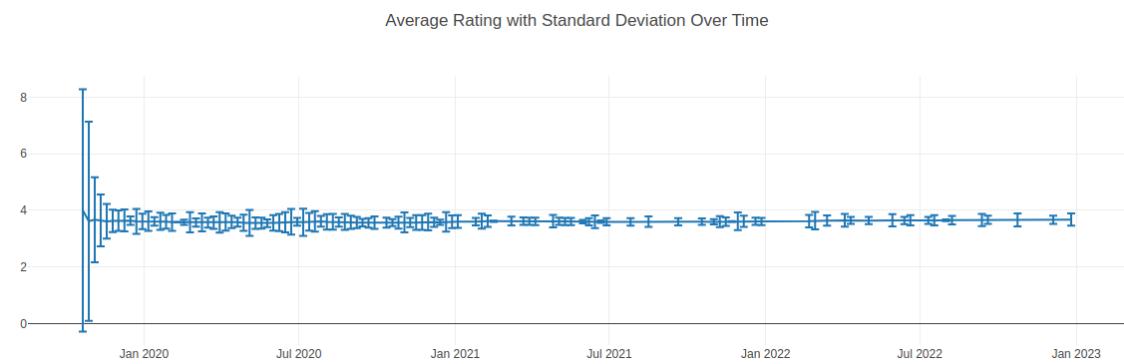


Figura B.2: Média de Pontuação com Desvio Padrão ao longo do tempo dos usuários.

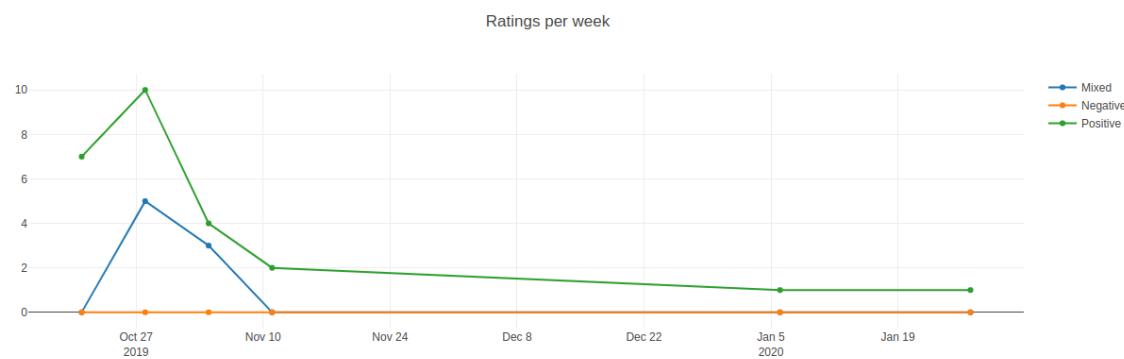


Figura B.3: Avaliações por semana dos especialistas.

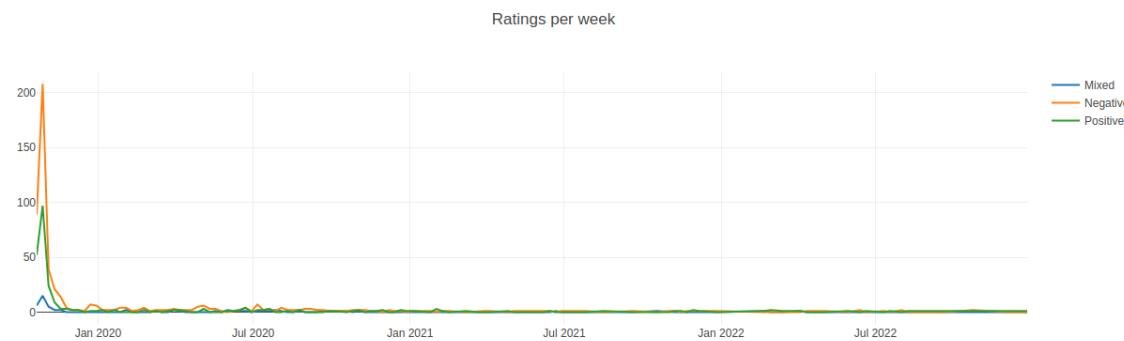


Figura B.4: Avaliações por semana dos usuários.

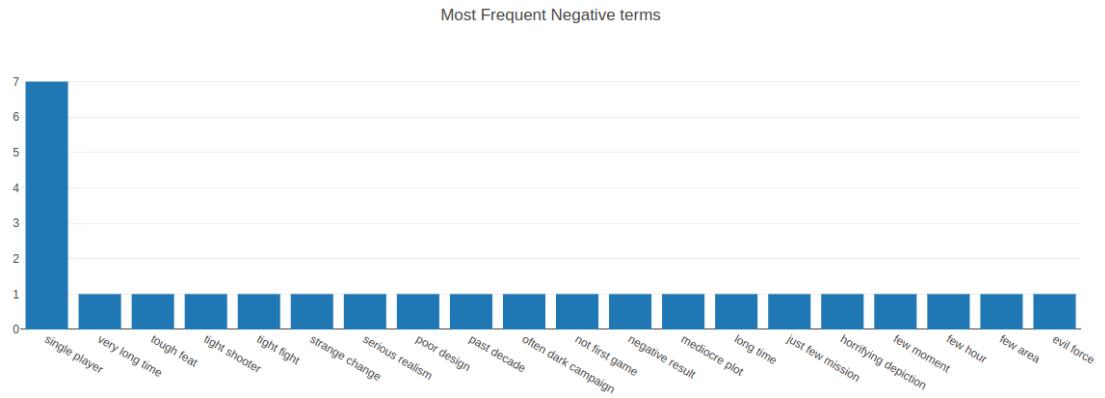


Figura B.5: Top termos negativos mais frequentes dos especialistas.

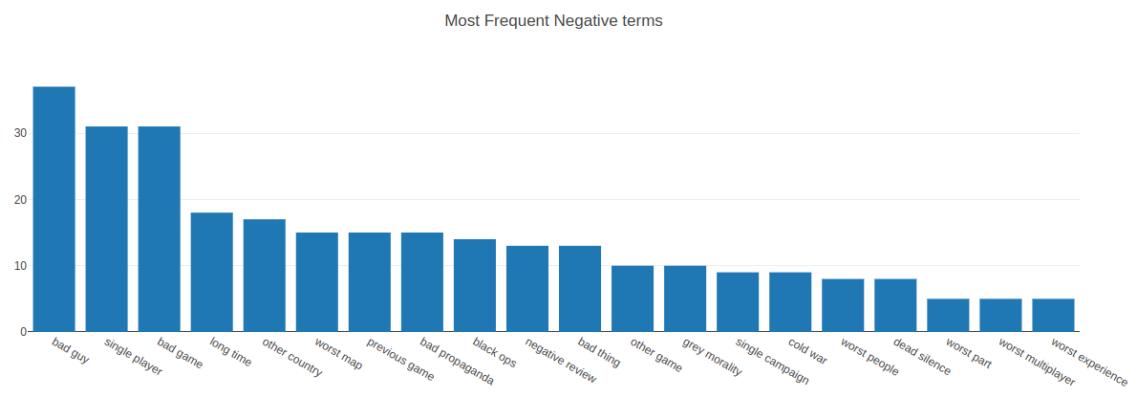


Figura B.6: Top termos negativos mais frequentes dos usuários.

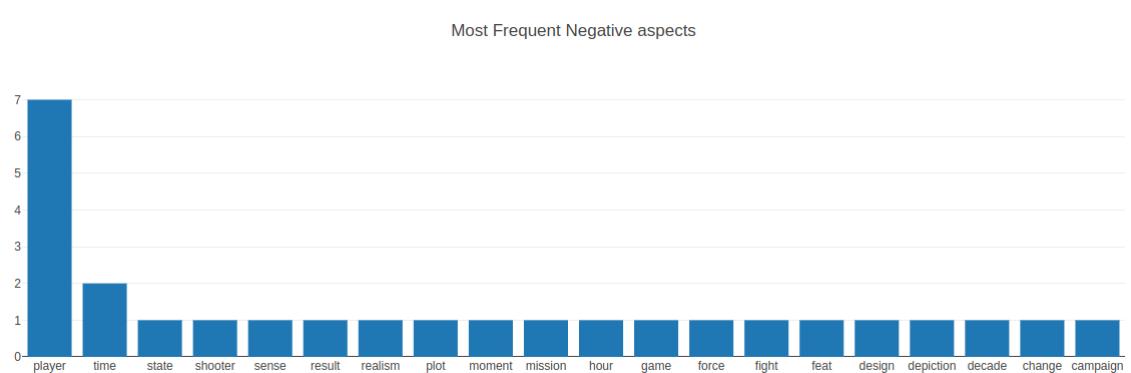


Figura B.7: Top aspectos negativos mais frequentes dos especialistas.

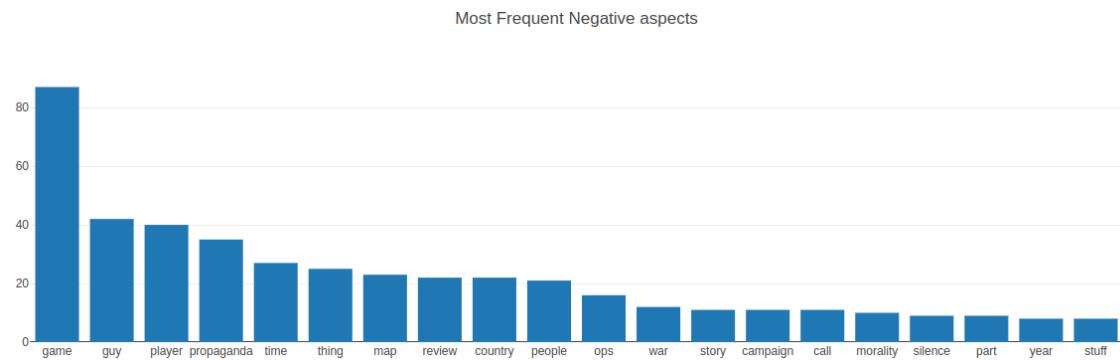


Figura B.8: Top aspectos negativos mais frequentes dos usuários.

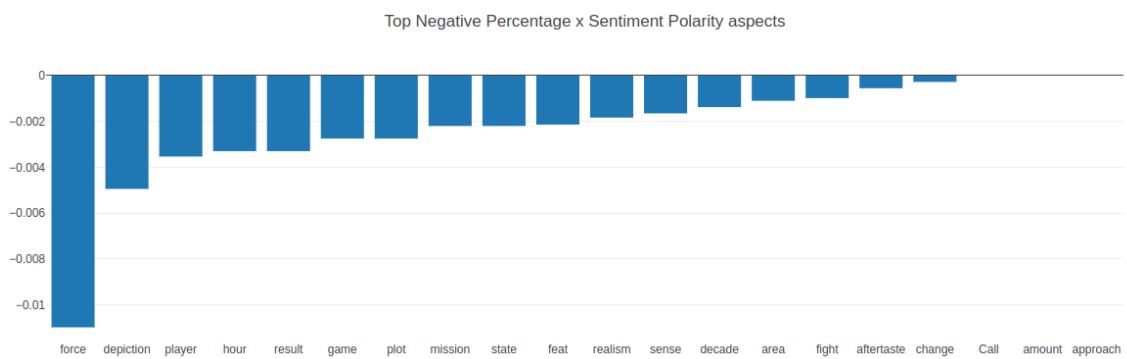


Figura B.9: Top aspectos negativos com maior valor porcentagem vezes polaridade dos especialistas.

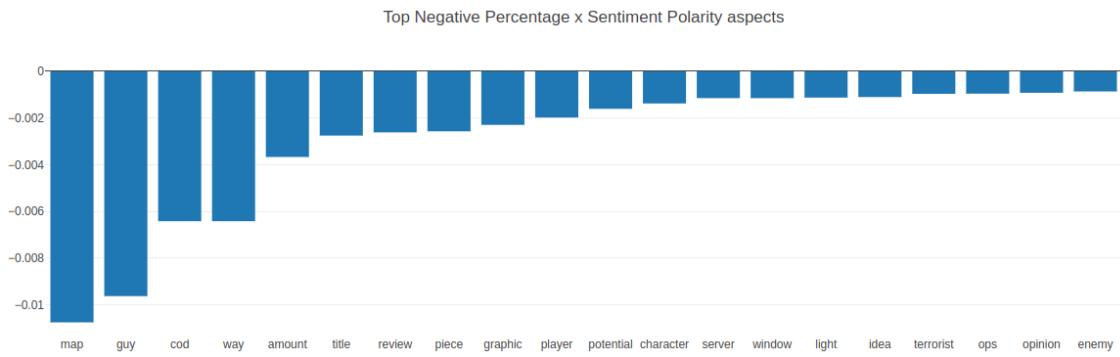


Figura B.10: Top aspectos negativos com maior valor porcentagem vezes polaridade dos usuários.

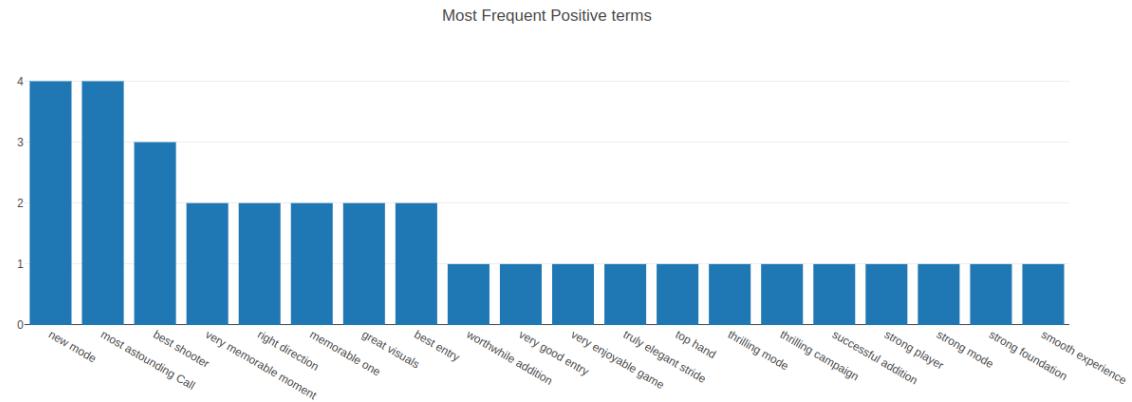


Figura B.11: Top termos positivos mais frequentes dos especialistas.

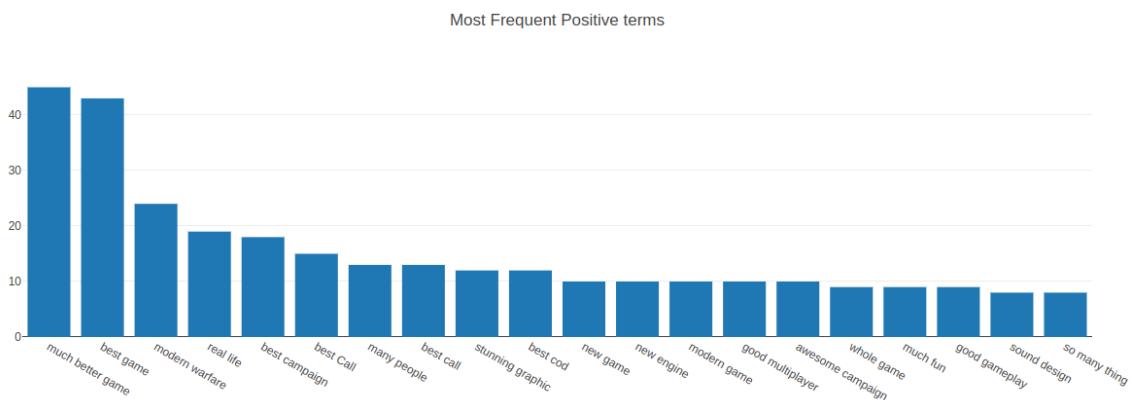


Figura B.12: Top termos positivos mais frequentes dos usuários.

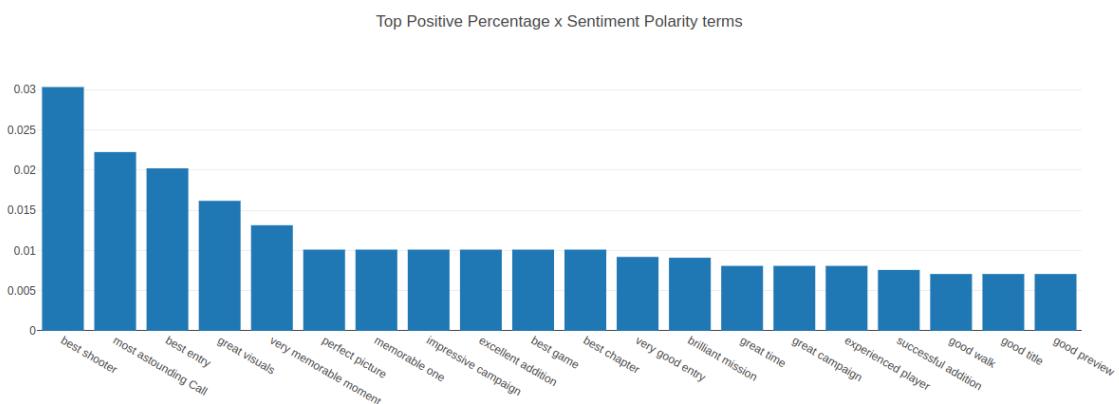


Figura B.13: Top termos positivos com maior valor porcentagem vezes polaridade dos especialistas.

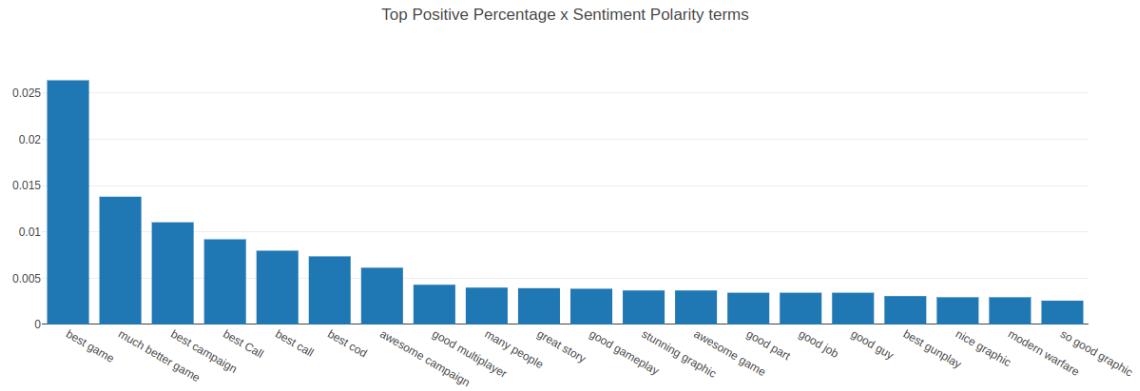


Figura B.14: Top termos positivos com maior valor porcentagem vezes polaridade dos usuários.

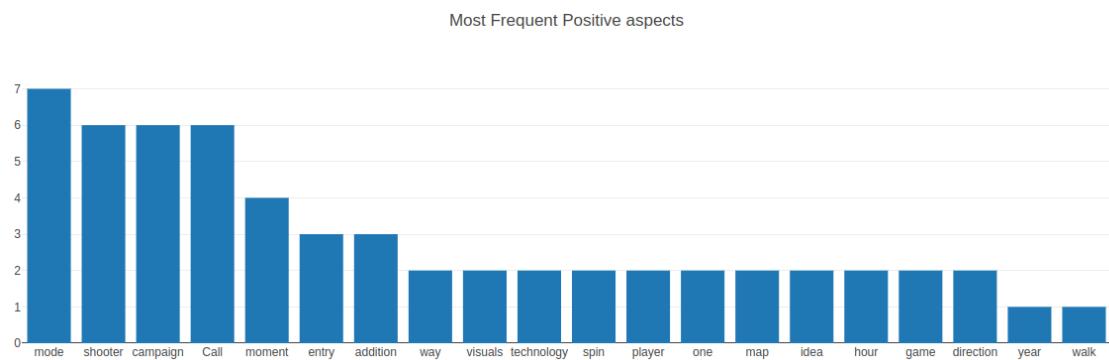


Figura B.15: Top aspectos positivos mais frequentes dos especialistas.

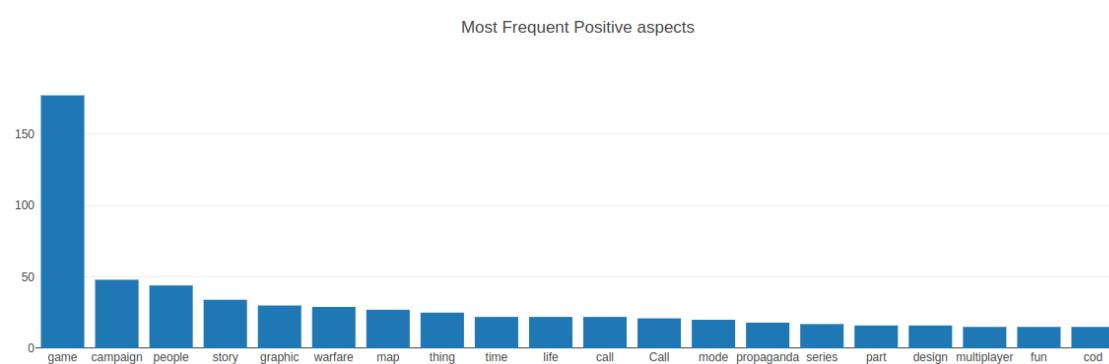


Figura B.16: Top aspectos positivos mais frequentes dos usuários.

Top Positive Percentage x Sentiment Polarity aspects

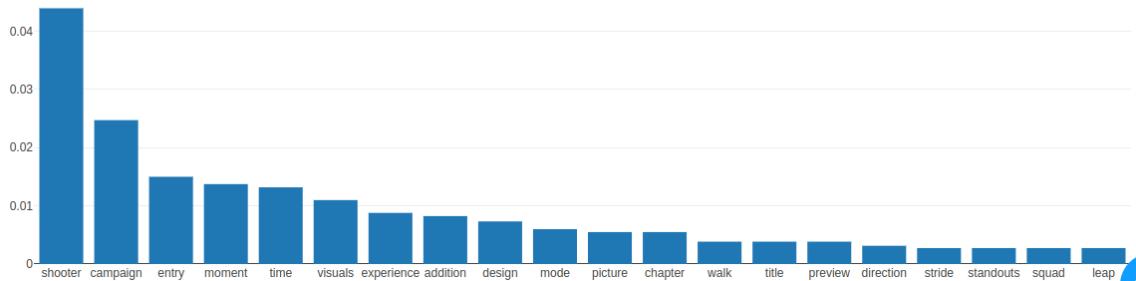


Figura B.17: Top aspectos positivos com maior valor porcentagem vezes polaridade dos especialistas.

Top Positive Percentage x Sentiment Polarity aspects

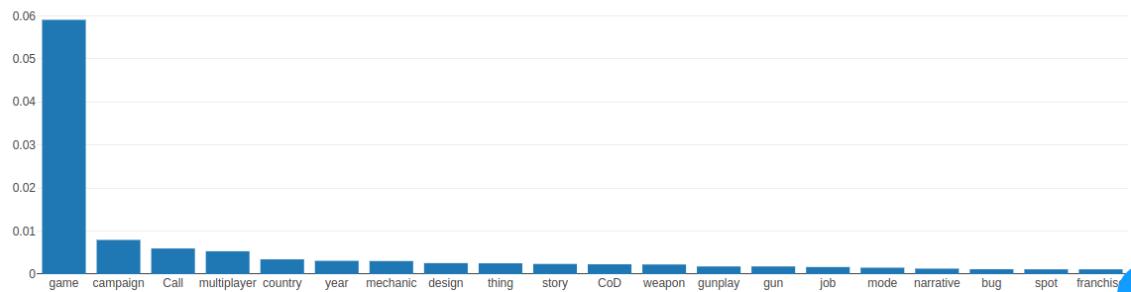


Figura B.18: Top aspectos positivos com maior valor porcentagem vezes polaridade dos usuários.

Apêndice C

Mais Gráficos e Diagramas gerados no Experimento 3

Este apêndice tem como objetivo apresentar todos os gráficos e diagramas gerados no Experimento 3 em que foi usado o conjunto de dados do jogo *Elden Ring* da plataforma *PC*, porém não foram utilizados na análise e avaliação dos resultados finais no Capítulo 4.

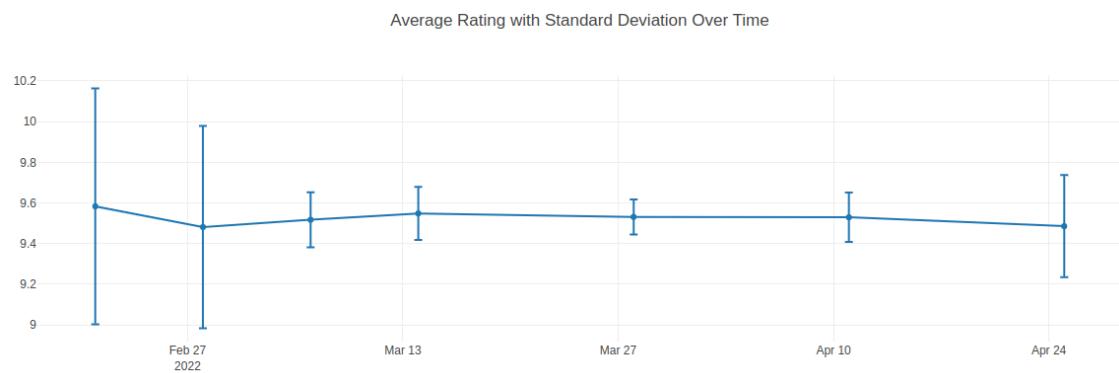


Figura C.1: Média de Pontuação com Desvio Padrão ao longo do tempo dos especialistas.

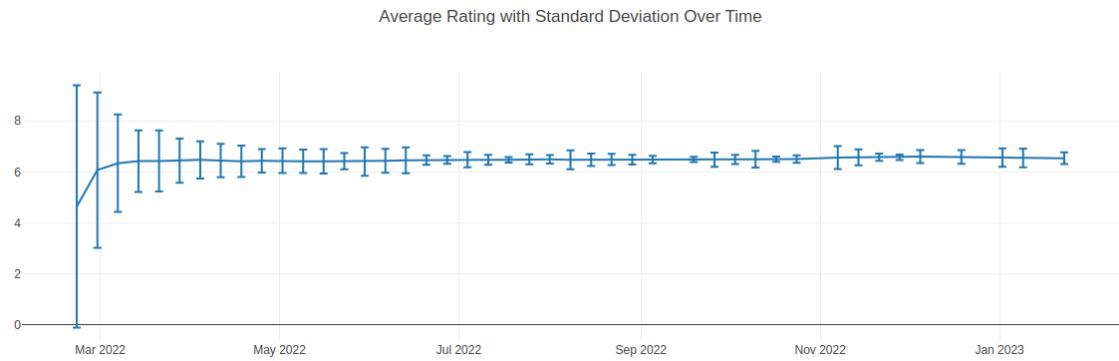


Figura C.2: Média de Pontuação com Desvio Padrão ao longo do tempo dos usuários.

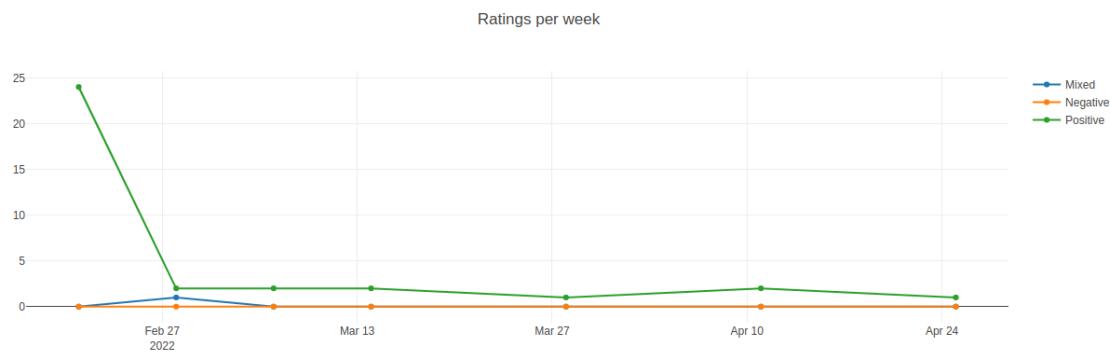


Figura C.3: Avaliações por semana dos especialistas.

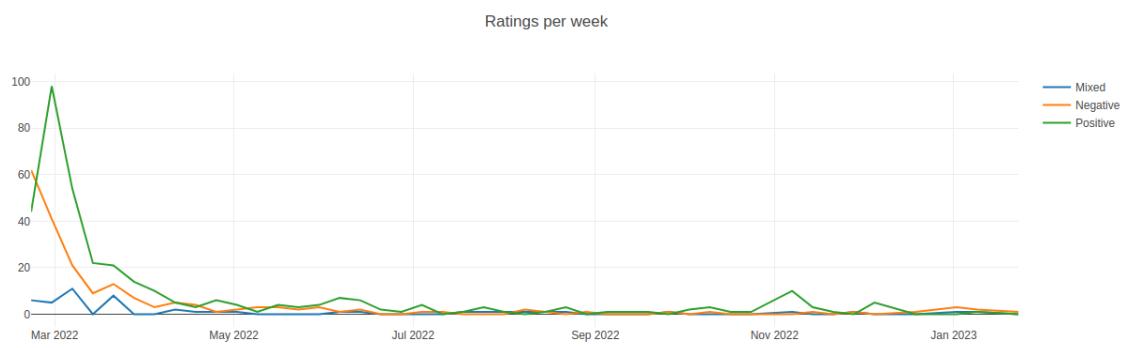


Figura C.4: Avaliações por semana dos usuários.

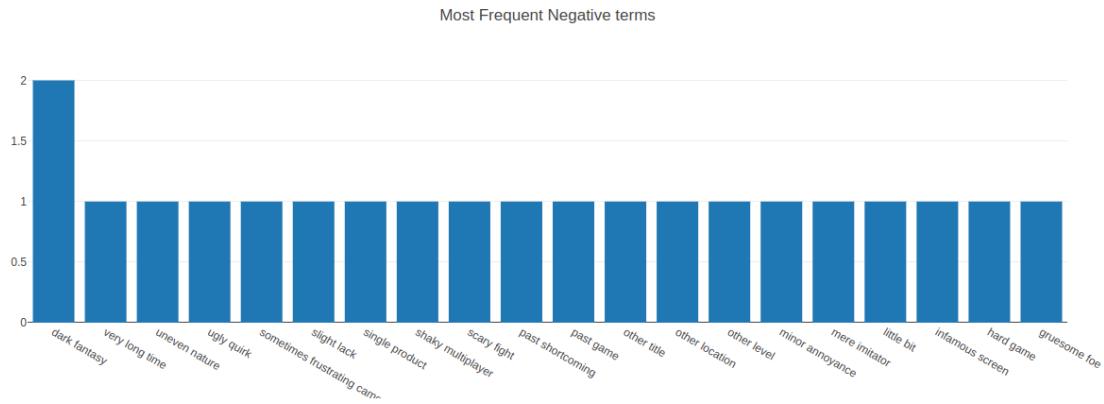


Figura C.5: Top termos negativos mais frequentes dos especialistas.

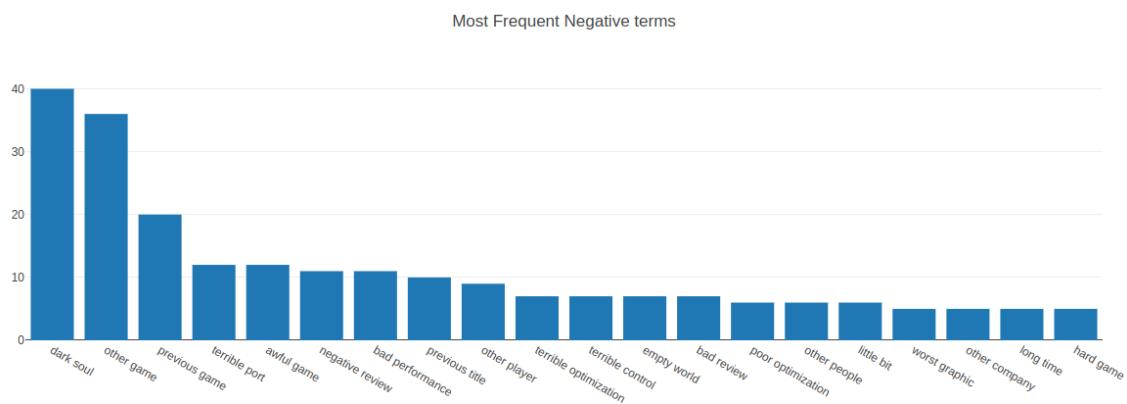


Figura C.6: Top termos negativos mais frequentes dos usuários.

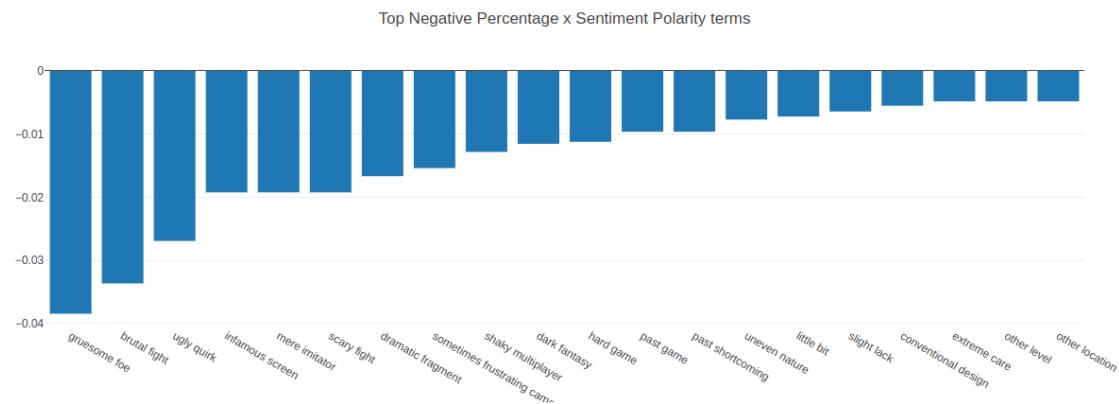


Figura C.7: Top termos negativos com maior valor porcentagem vezes polaridade dos especialistas.

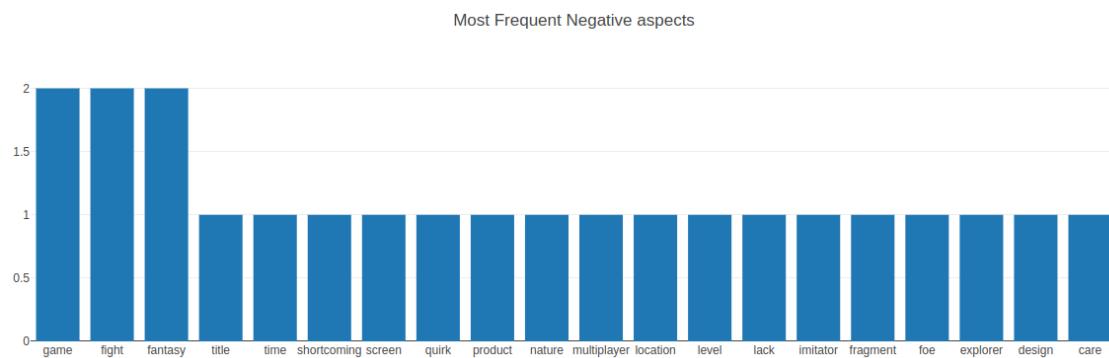


Figura C.8: Top aspectos negativos mais frequentes dos especialistas.

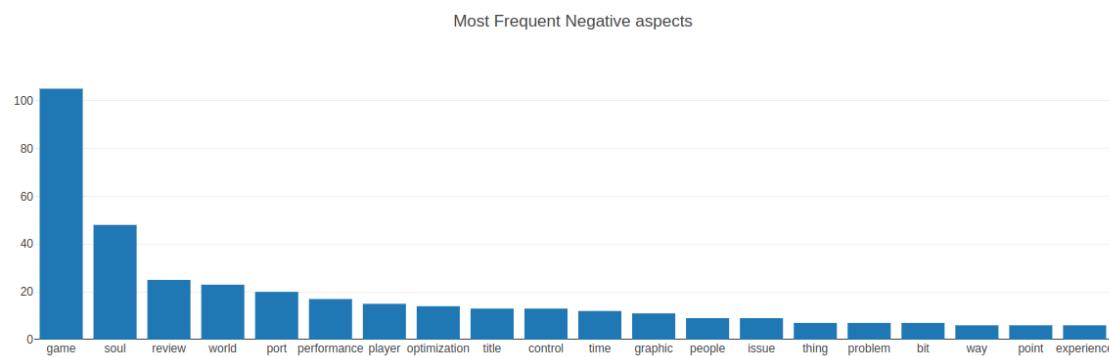


Figura C.9: Top aspectos negativos mais frequentes dos usuários.

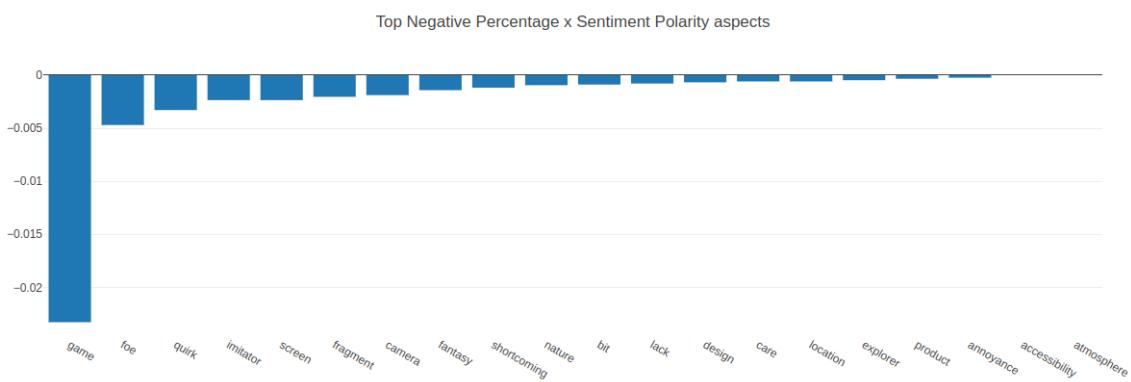


Figura C.10: Top aspectos negativos com maior valor porcentagem vezes polaridade dos especialistas.

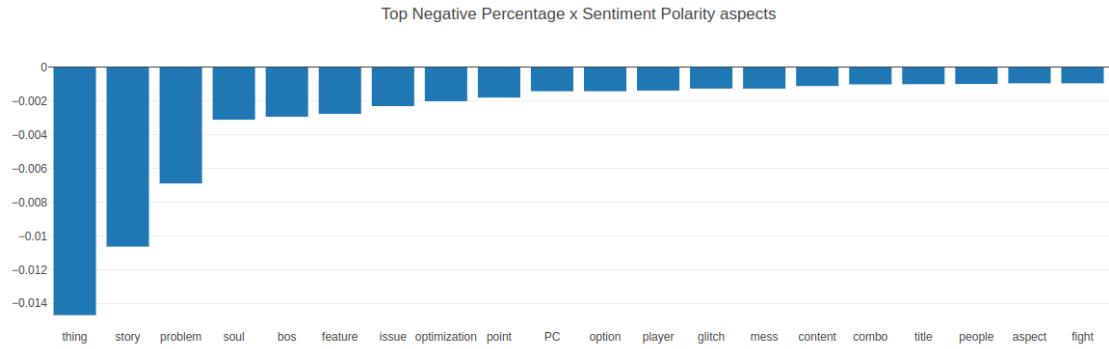


Figura C.11: Top aspectos negativos com maior valor porcentagem vezes polaridade dos usuários.

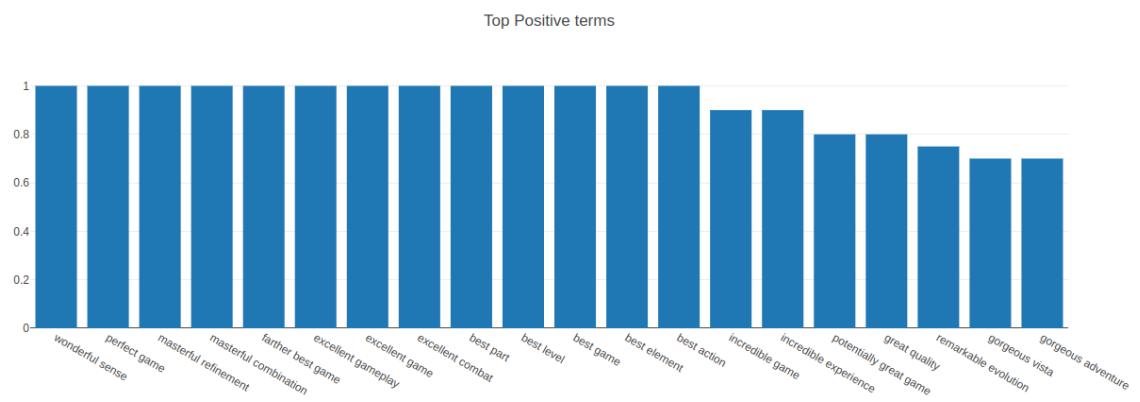


Figura C.12: Top termos positivos dos especialistas.

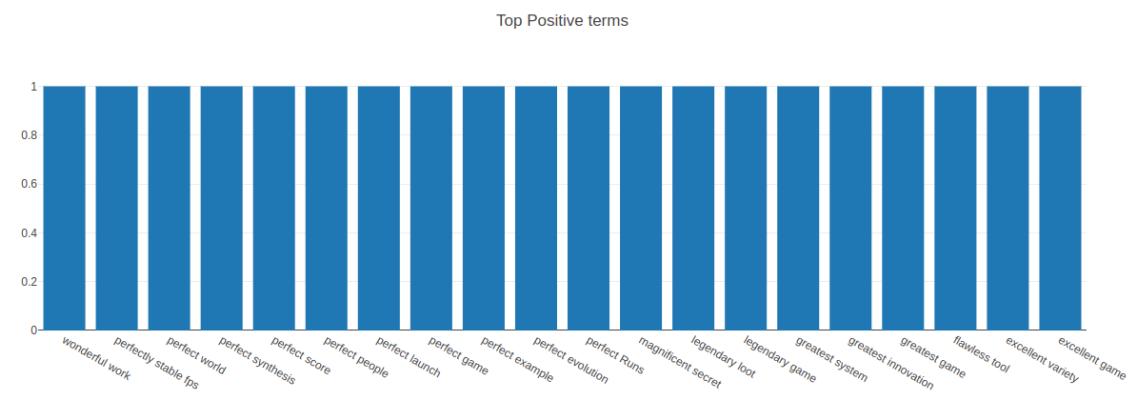


Figura C.13: Top termos positivos dos usuários.

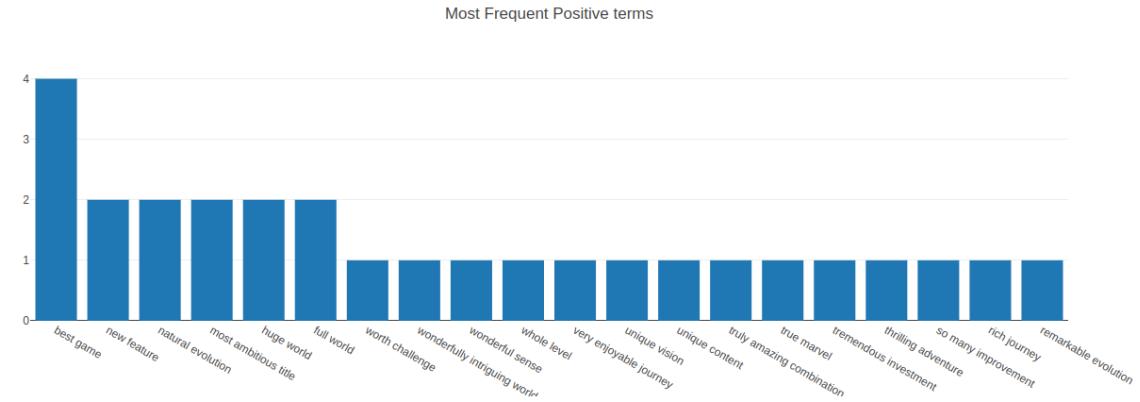


Figura C.14: Top termos positivos mais frequentes dos especialistas.

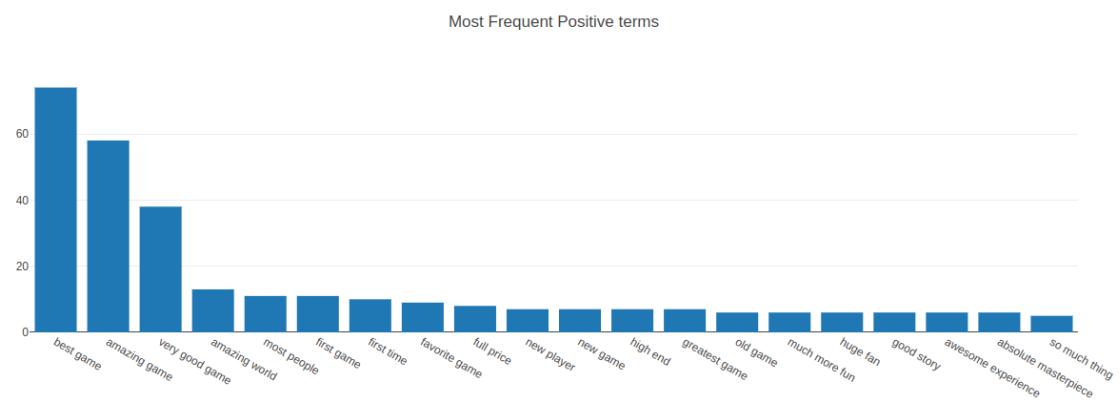


Figura C.15: Top termos positivos mais frequentes dos usuários.

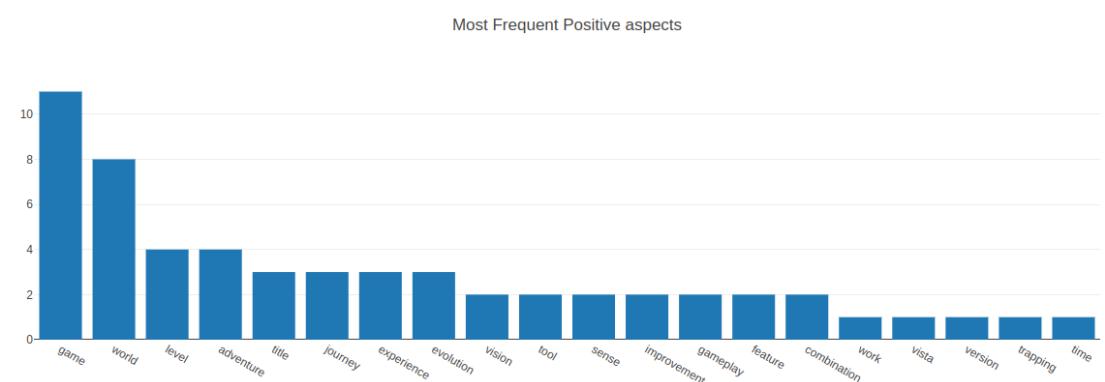


Figura C.16: Top aspectos positivos mais frequentes dos especialistas.

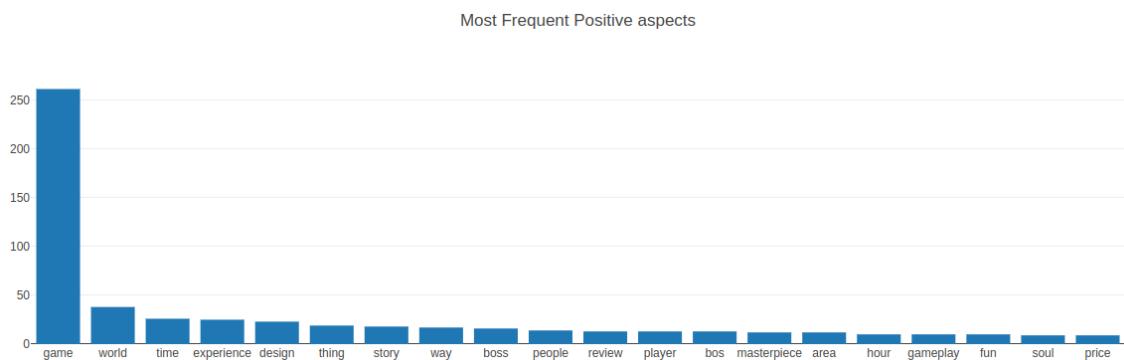


Figura C.17: Top aspectos positivos mais frequentes dos usuários.

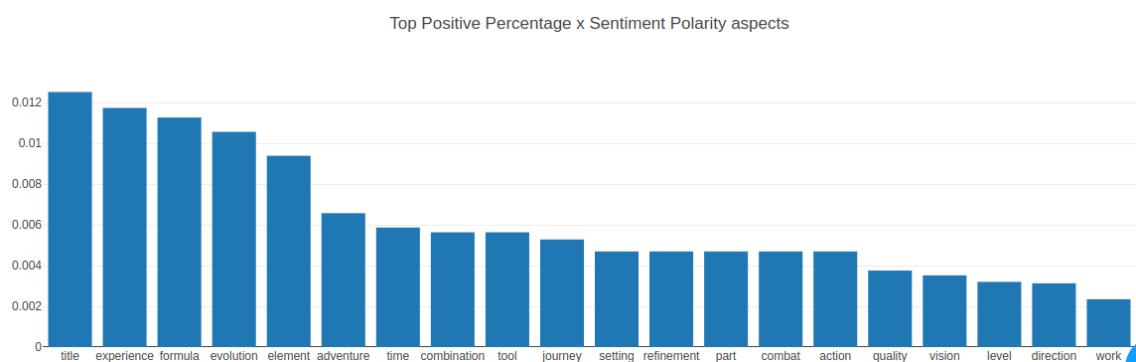


Figura C.18: Top aspectos positivos com maior valor porcentagem vezes polaridade dos especialistas.

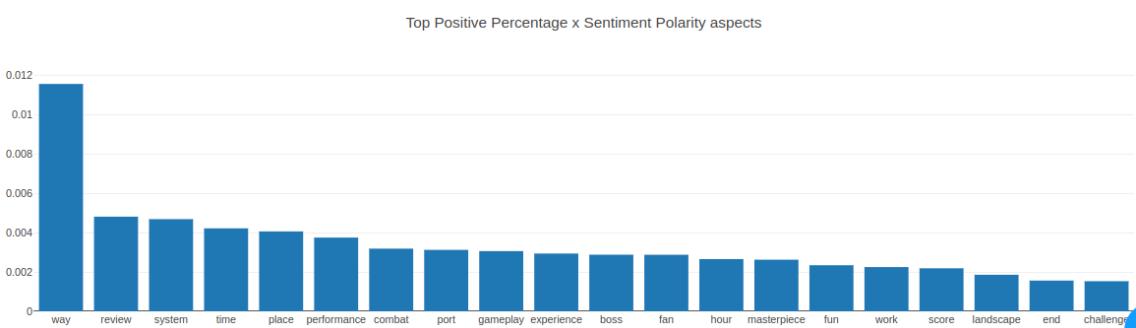


Figura C.19: Top aspectos positivos com maior valor porcentagem vezes polaridade dos usuários.

Apêndice D

Mais Gráficos e Diagramas gerados no Experimento 4

Este apêndice tem como objetivo apresentar todos os gráficos e diagramas gerados no Experimento 4 em que foi usado o conjunto de dados do jogo *Cyberpunk 2077* da plataforma *PC*, porém não foram utilizados na análise e avaliação dos resultados finais no Capítulo 4.

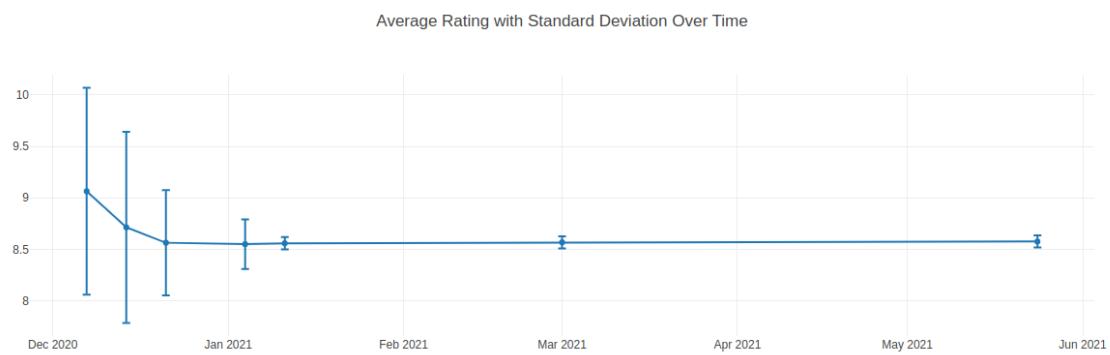


Figura D.1: Média de Pontuação com Desvio Padrão ao longo do tempo dos especialistas.

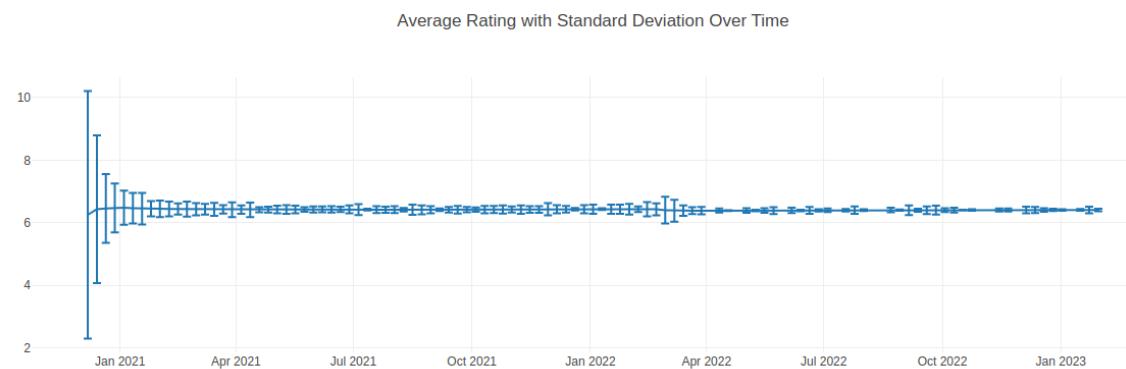


Figura D.2: Média de Pontuação com Desvio Padrão ao longo do tempo dos usuários.

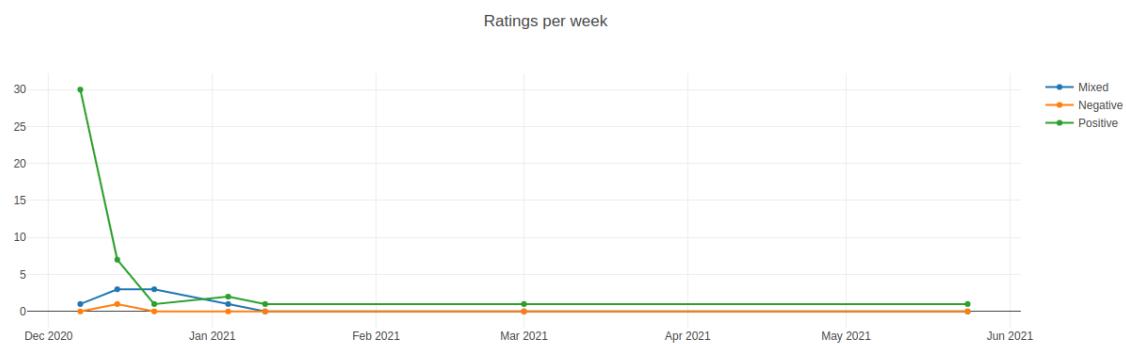


Figura D.3: Avaliações por semana dos especialistas.

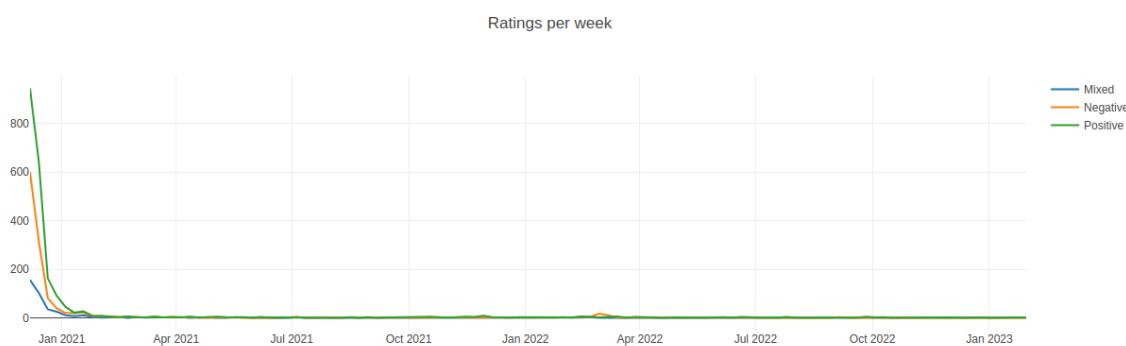


Figura D.4: Avaliações por semana dos usuários.

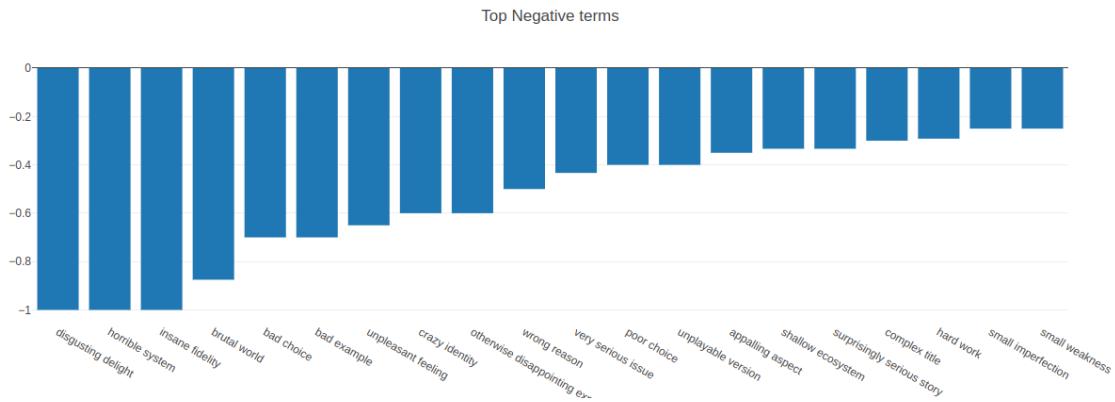


Figura D.5: Top termos negativos dos especialistas.

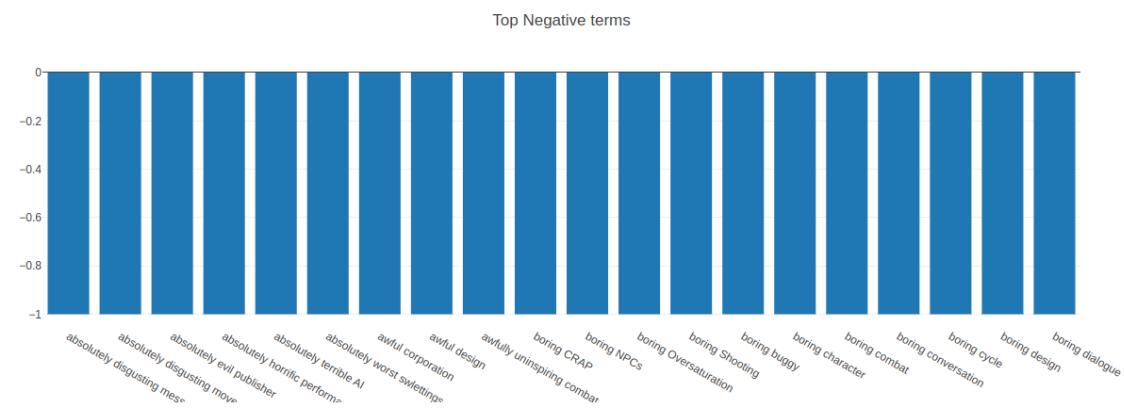


Figura D.6: Top termos negativos dos usuários.

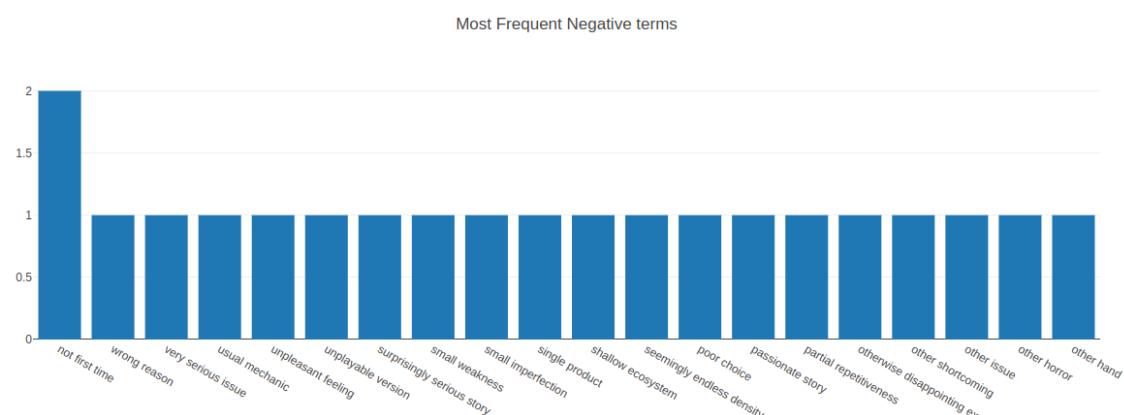


Figura D.7: Top termos negativos mais frequentes dos especialistas.

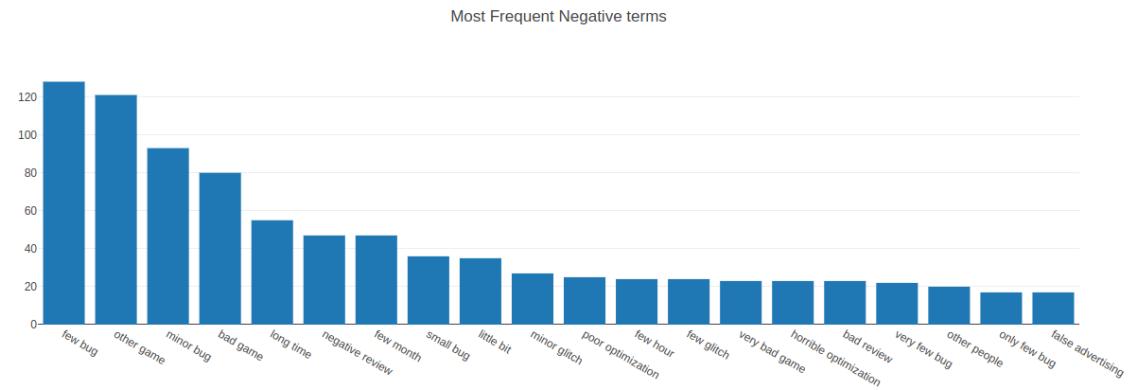


Figura D.8: Top termos negativos mais frequentes dos usuários.

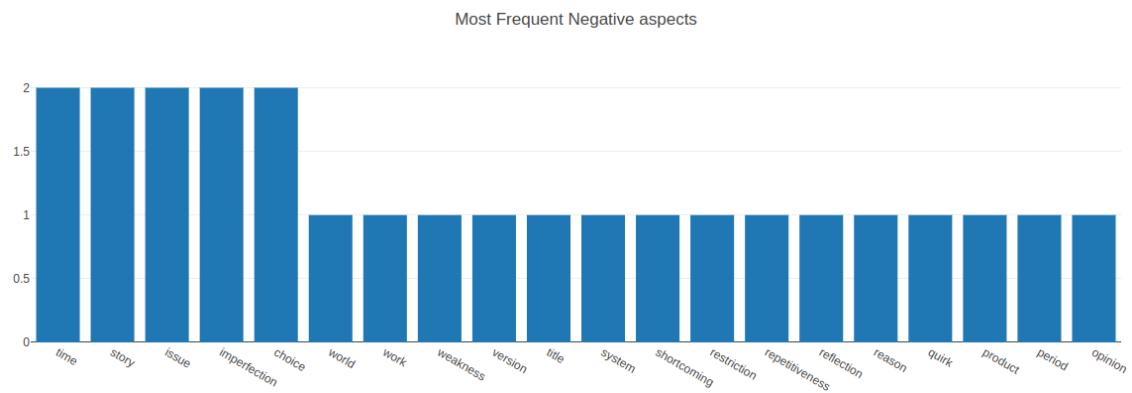


Figura D.9: Top aspectos negativos mais frequentes dos especialistas.

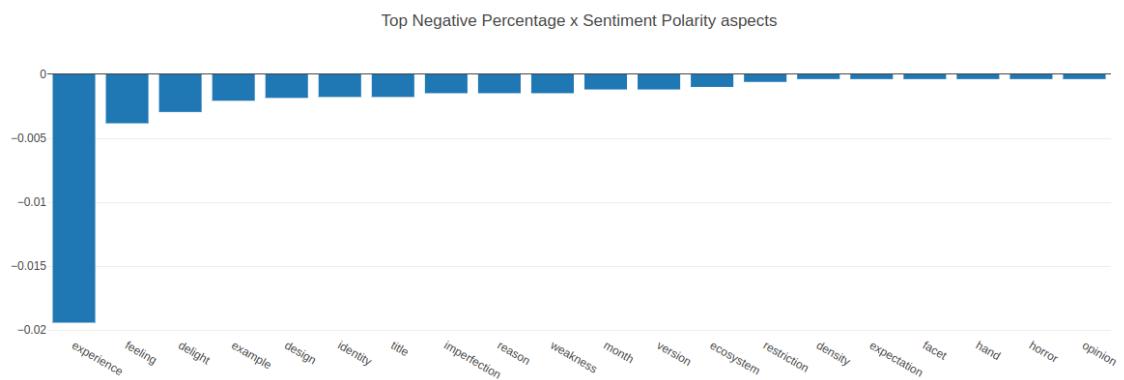


Figura D.10: Top aspectos negativos com maior valor porcentagem vezes polaridade dos especialistas.

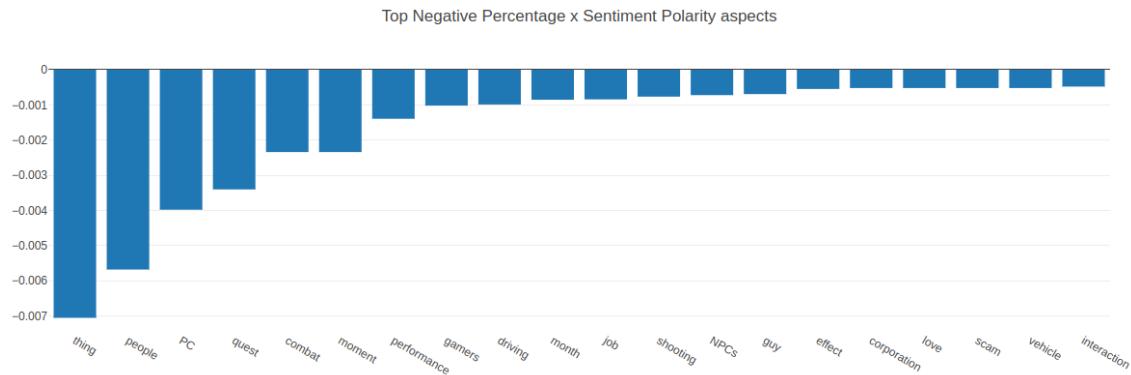


Figura D.11: Top aspectos negativos com maior valor porcentagem vezes polaridade dos usuários.

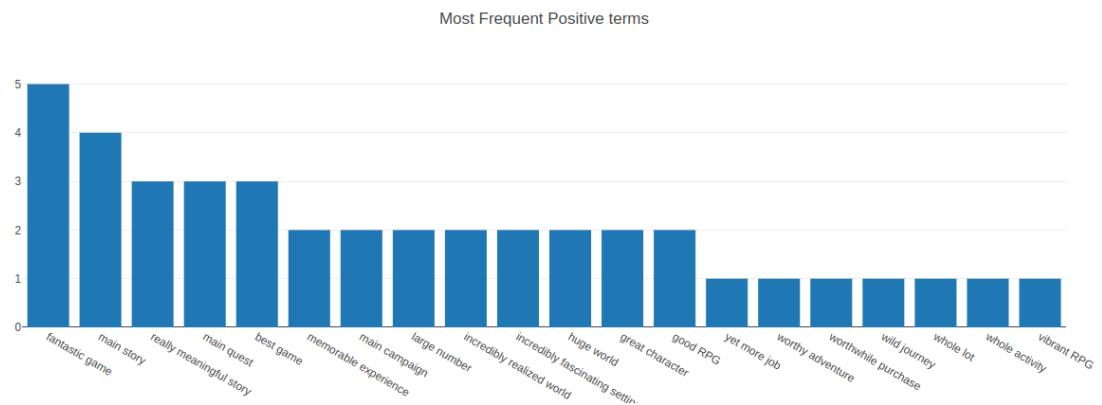


Figura D.12: Top termos positivos mais frequentes dos especialistas.

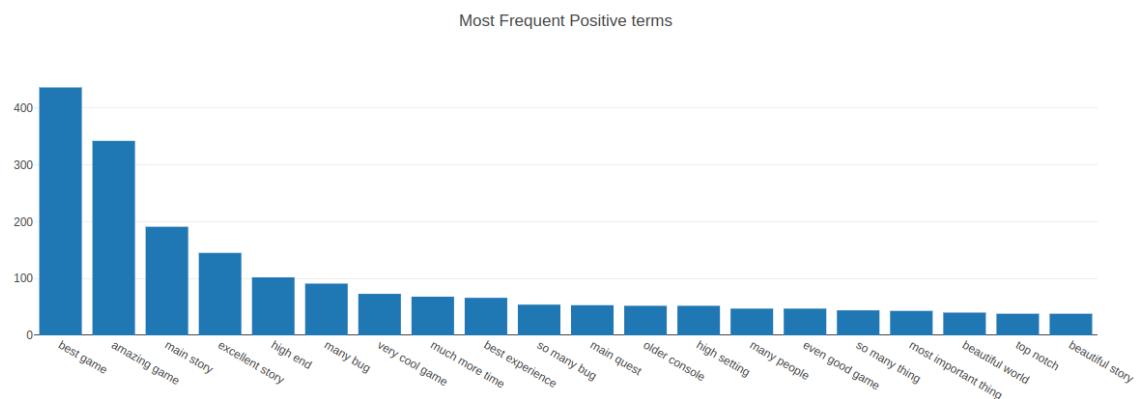


Figura D.13: Top termos positivos mais frequentes dos usuários.

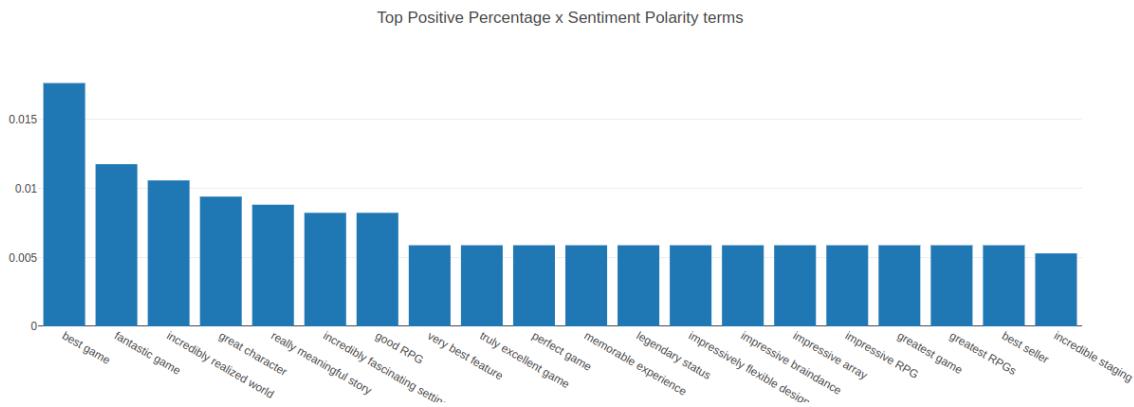


Figura D.14: Top termos positivos com maior valor porcentagem vezes polaridade dos especialistas.

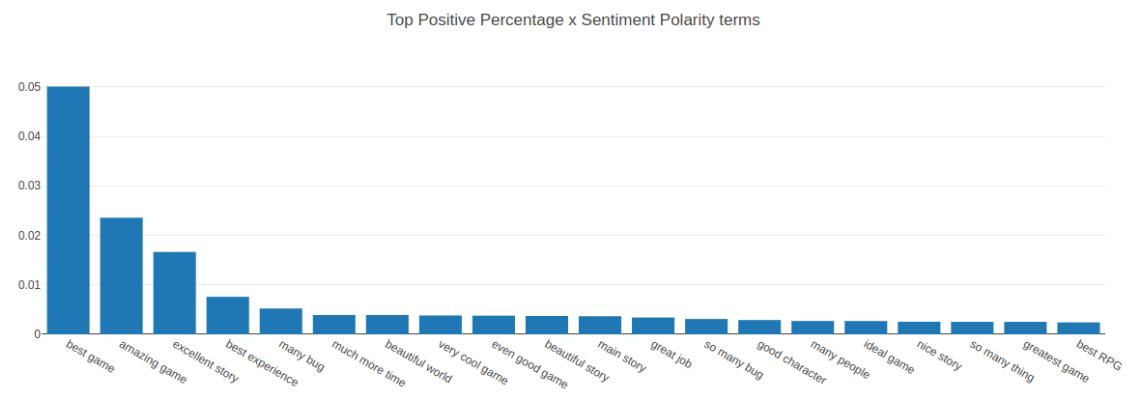


Figura D.15: Top termos positivos com maior valor porcentagem vezes polaridade dos usuários.

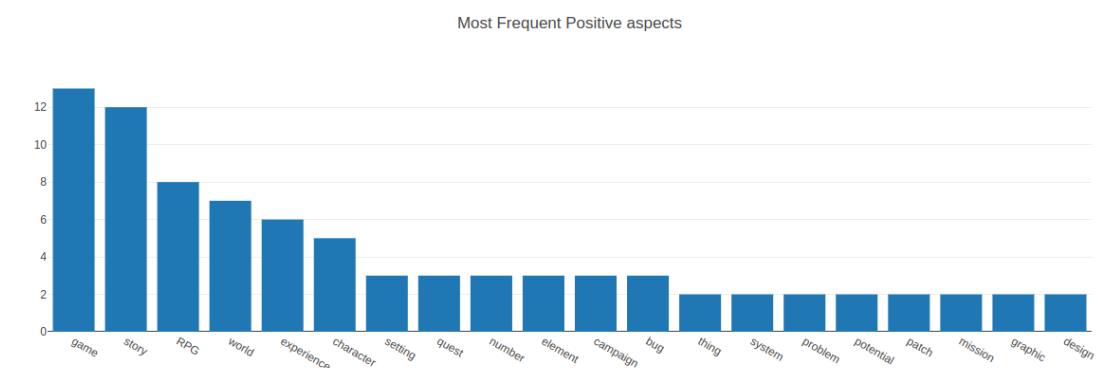


Figura D.16: Top aspectos positivos mais frequentes dos especialistas.

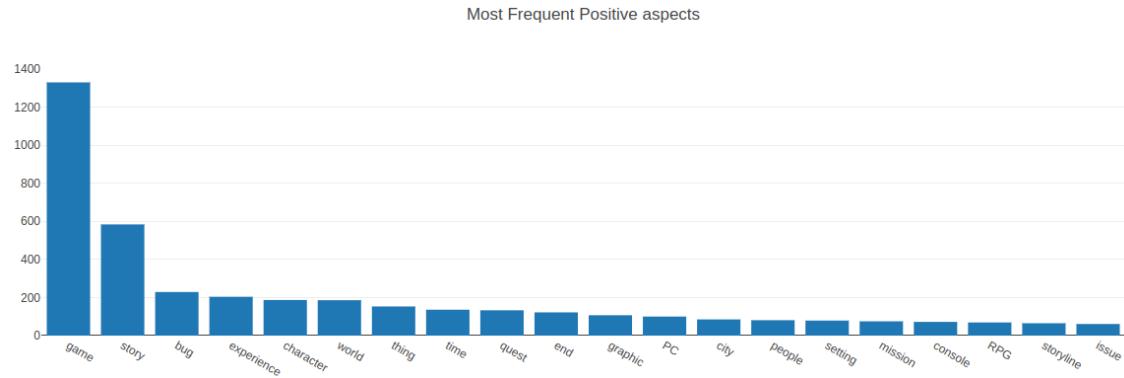


Figura D.17: Top aspectos positivos mais frequentes dos usuários.

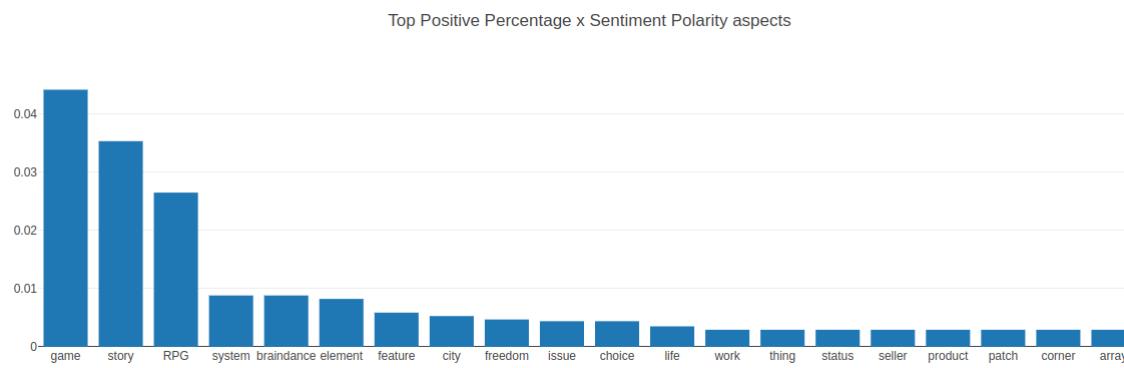


Figura D.18: Top aspectos positivos com maior valor porcentagem vezes polaridade dos especialistas.

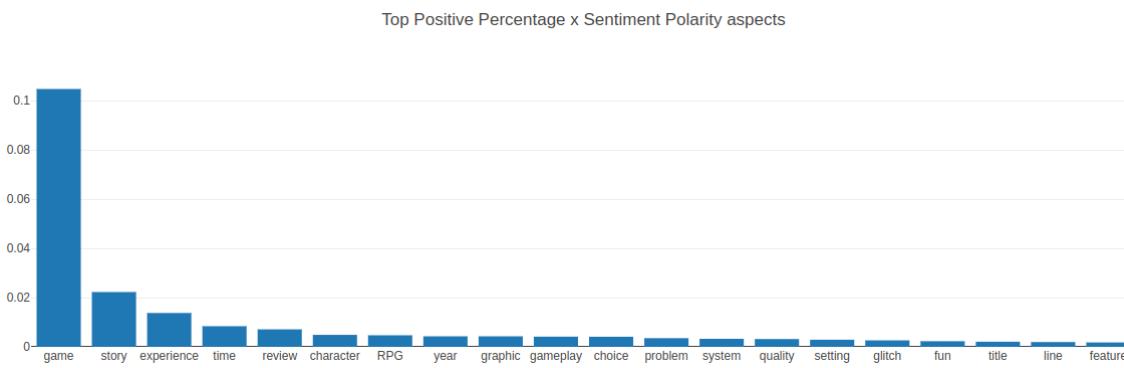


Figura D.19: Top aspectos positivos com maior valor porcentagem vezes polaridade dos usuários.