

# SteamDE: análises de *reviews* de jogos na Steam

Daniel de Viveiros Inácio  
Eduardo Darrazão



**STEAM®**



MINISTÉRIO DA  
**EDUCAÇÃO**



# Descrição dos Dados

- *Reviews* de jogadores na Steam de 48 jogos diferentes;
- Entre 20/12/2010 até 16/02/2019;
- Dados:
  - Data da postagem;
  - Título do jogo;
  - *Review* em si;
  - Se o autor do *review* recomenda ou não o jogo;
  - Quantas horas o autor jogou o jogo;
  - Quantas pessoas marcaram o *review* como engraçado;
  - Quantas pessoas marcaram o *review* como útil;
  - Se o jogo estava em *Early Access* ou não.

Disponível em:

<https://www.kaggle.com/datasets/luthfim/steam-reviews-dataset>



# Perguntas de Pesquisa

- Análise de sentimentos dos reviews, como se comportam dependendo se está recomendando ou não o jogo em questão, se tem relação com o total de horas jogadas;
- Analisar tópicos pertinentes aos jogos, sobre o que se trata, como evoluem ao longo do tempo;
- Investigar se os tópicos de cada review possuem relação com o tempo de jogo;
- Investigar se os tópicos dos reviews entre jogos com características similares possuem intersecção;
- Analisar se a aceitação dos jogos impacta na quantidade de horas jogadas.



# Limpeza e Processamento dos Dados

## Limpeza:

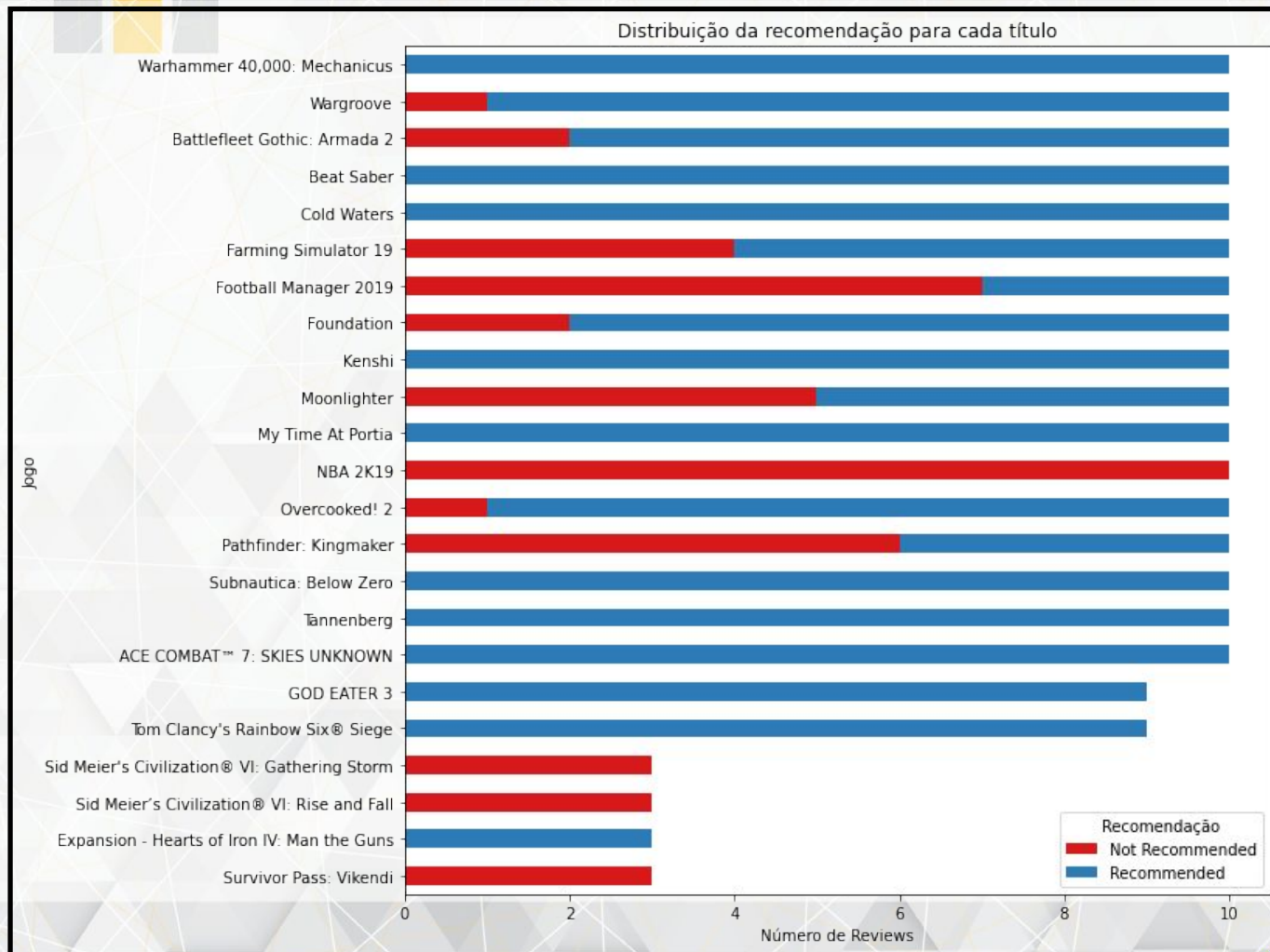
- Remoção de nulos e vazios;
- Filtro de idioma: FastText;
- Técnicas de processamento de texto:
  - Contrações, caixa baixa, pontuações, *tokenizar*, *stopwords*, *POS tags*, *lematizar*;
- Começo: 434.891 entradas. Pós limpeza: 399.085.

## Processamentos:

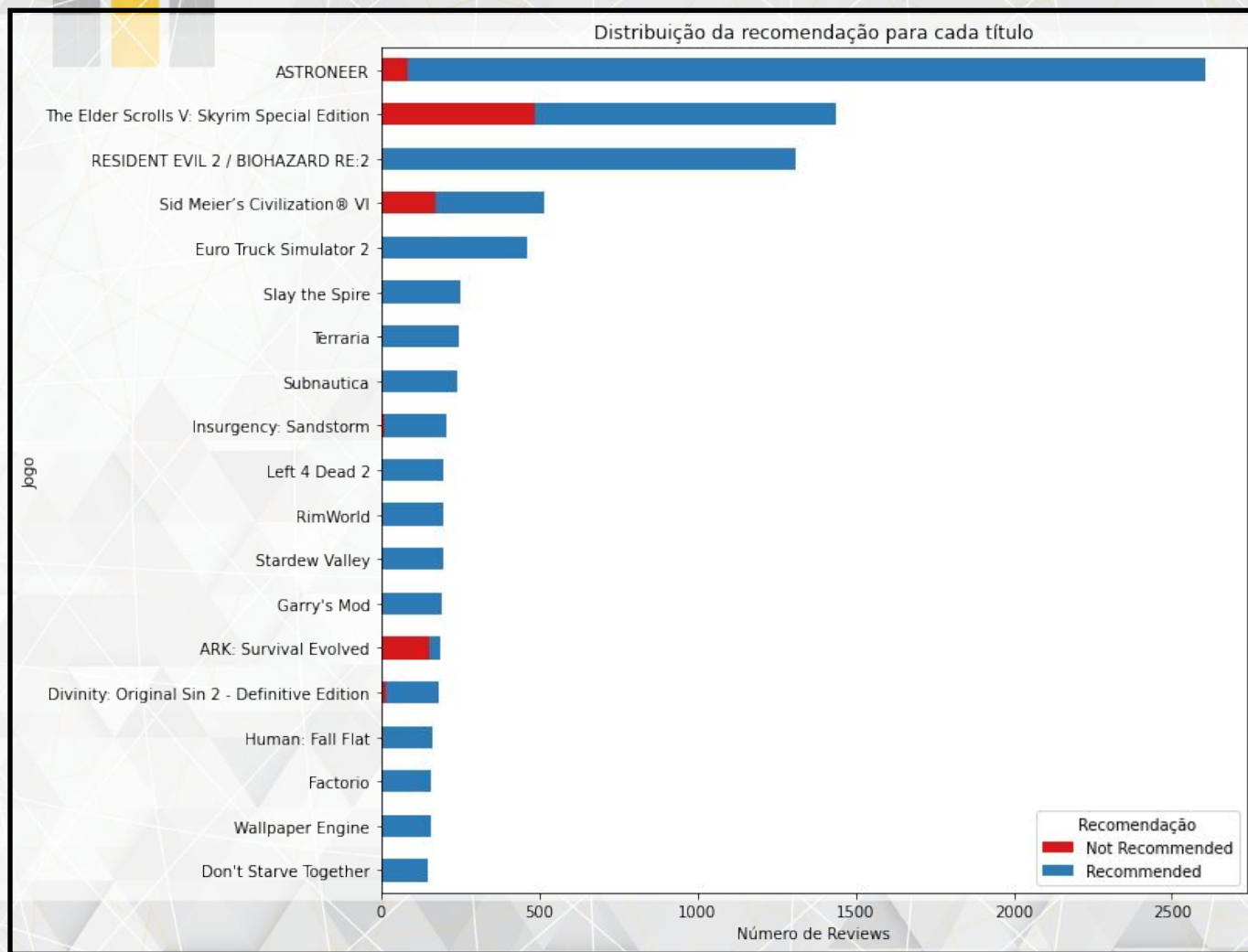
- Análise da distribuição e correlação dos dados;
- Análise de Sentimentos: TextBlob;
- Análise de tópicos: N-gramas e LSA.



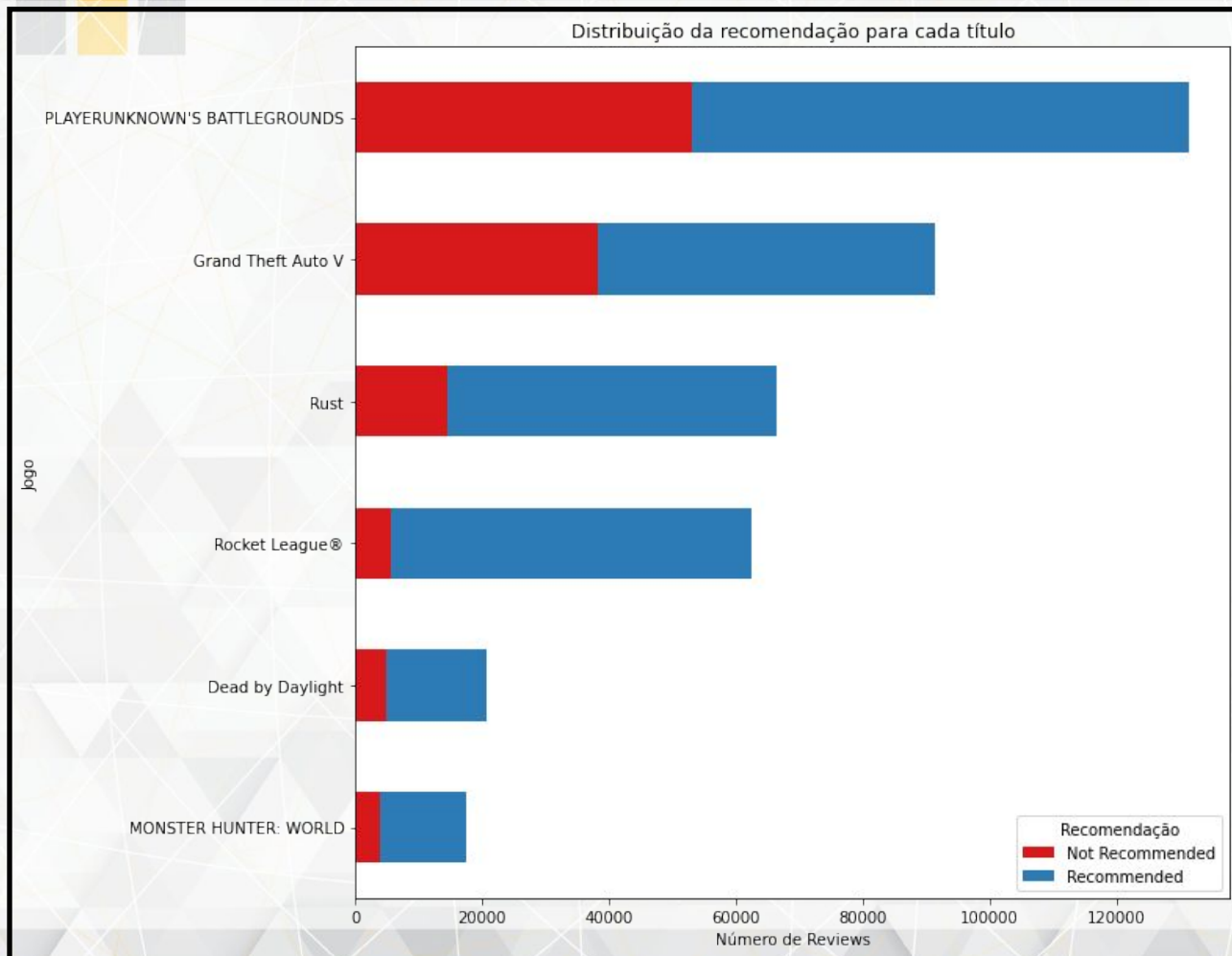
# Cobertura e Distribuição dos Dados



# Cobertura e Distribuição dos Dados



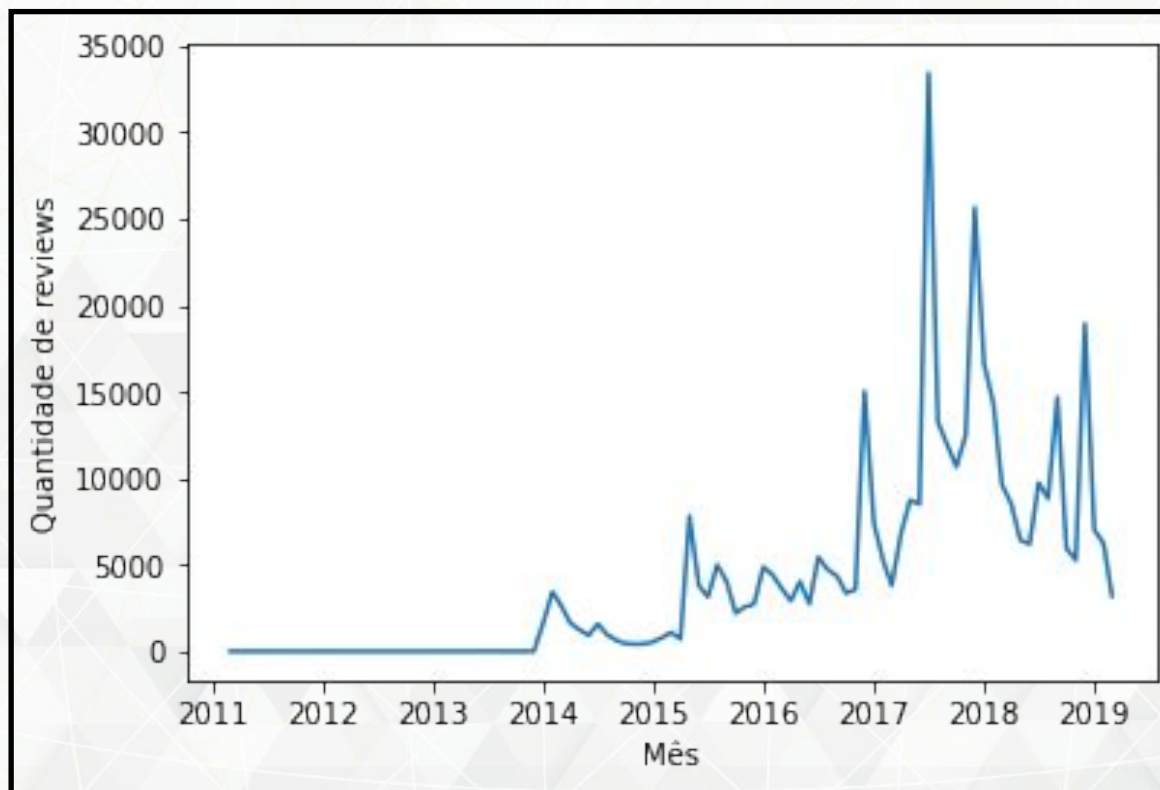
# Cobertura e Distribuição dos Dados





# Cobertura e Distribuição dos Dados

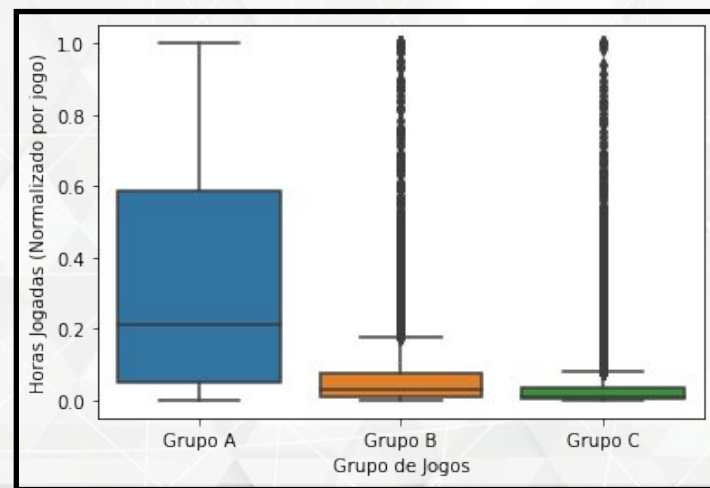
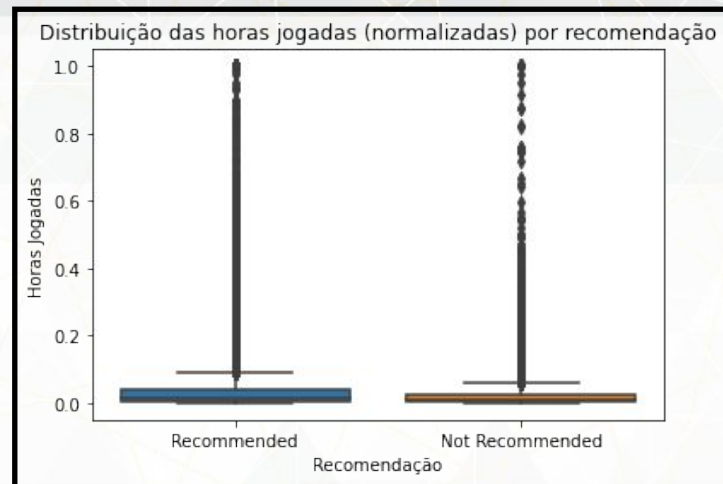
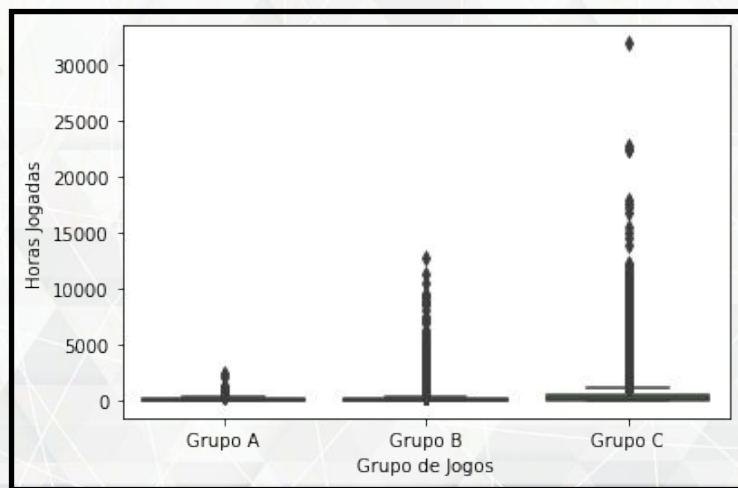
- Tópicos ao longo do tempo.





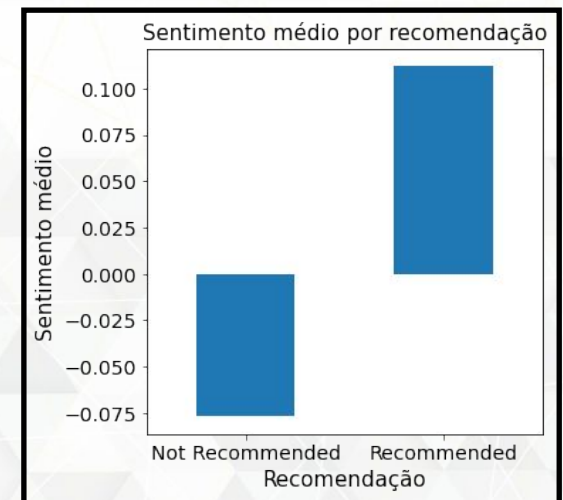
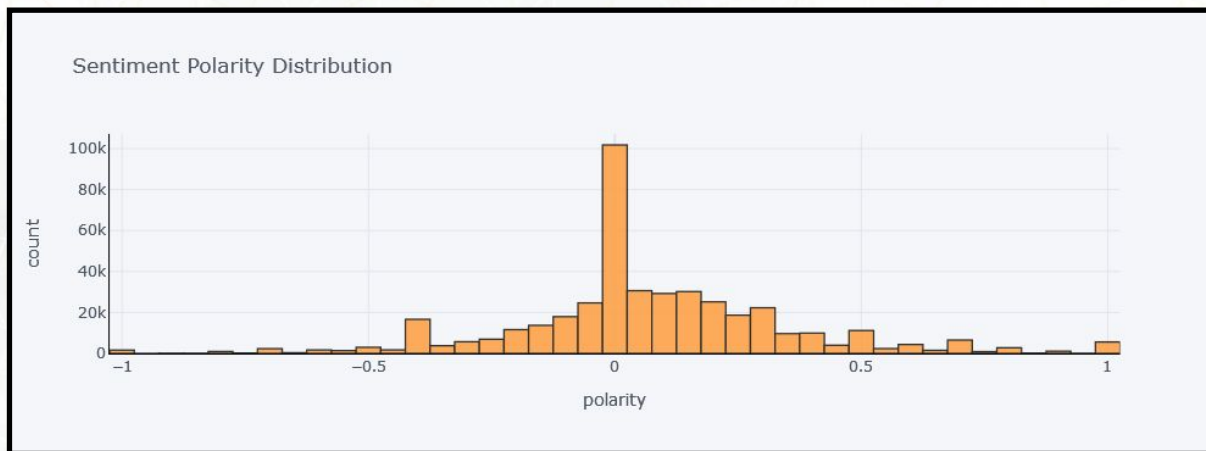
# Cobertura e Distribuição dos Dados

- Tópicos com tempo de jogo.



# Análise Exploratória

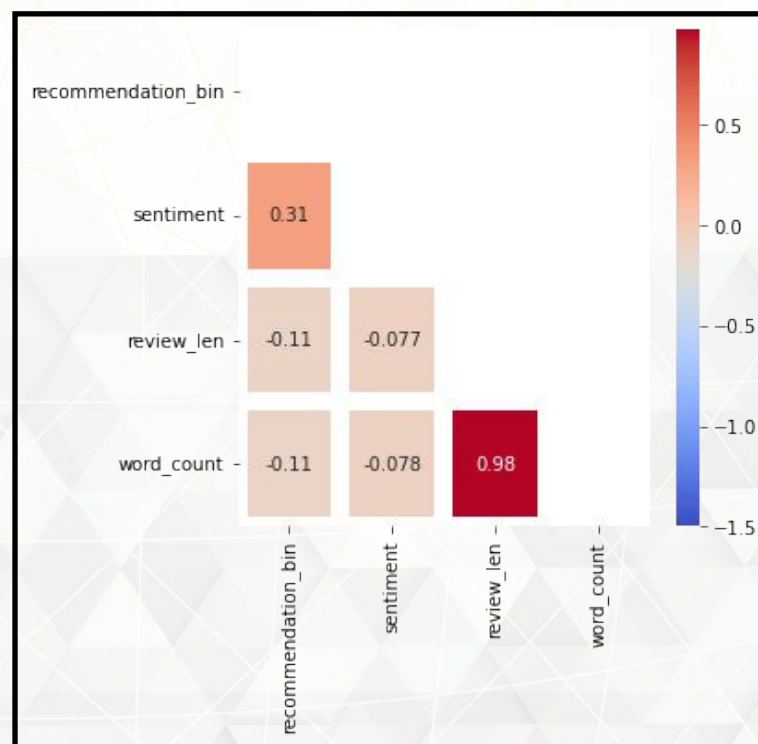
- Correlações com análise de sentimentos





# Análise Exploratória

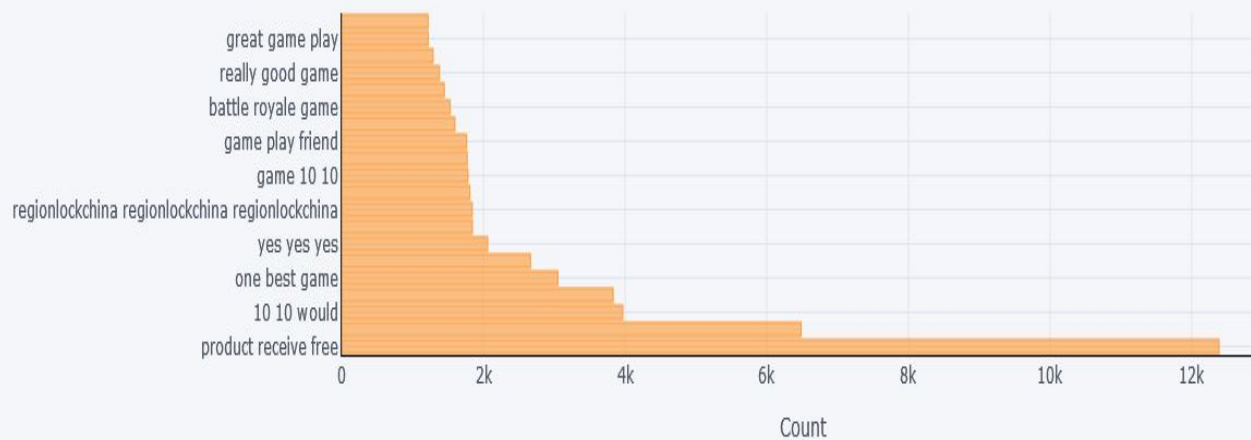
- Análise de sentimentos dos reviews, como se comportam dependendo se está recomendando ou não o jogo em questão.



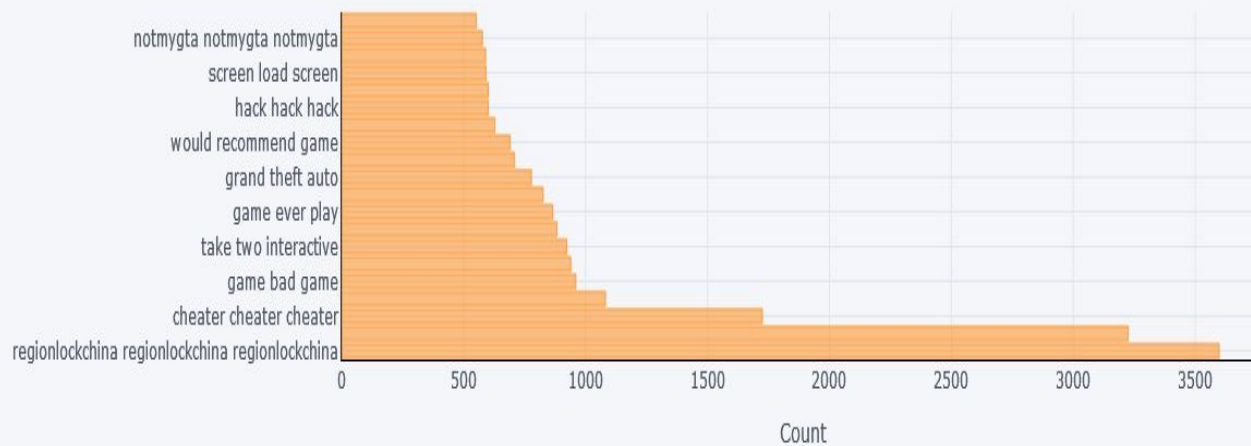


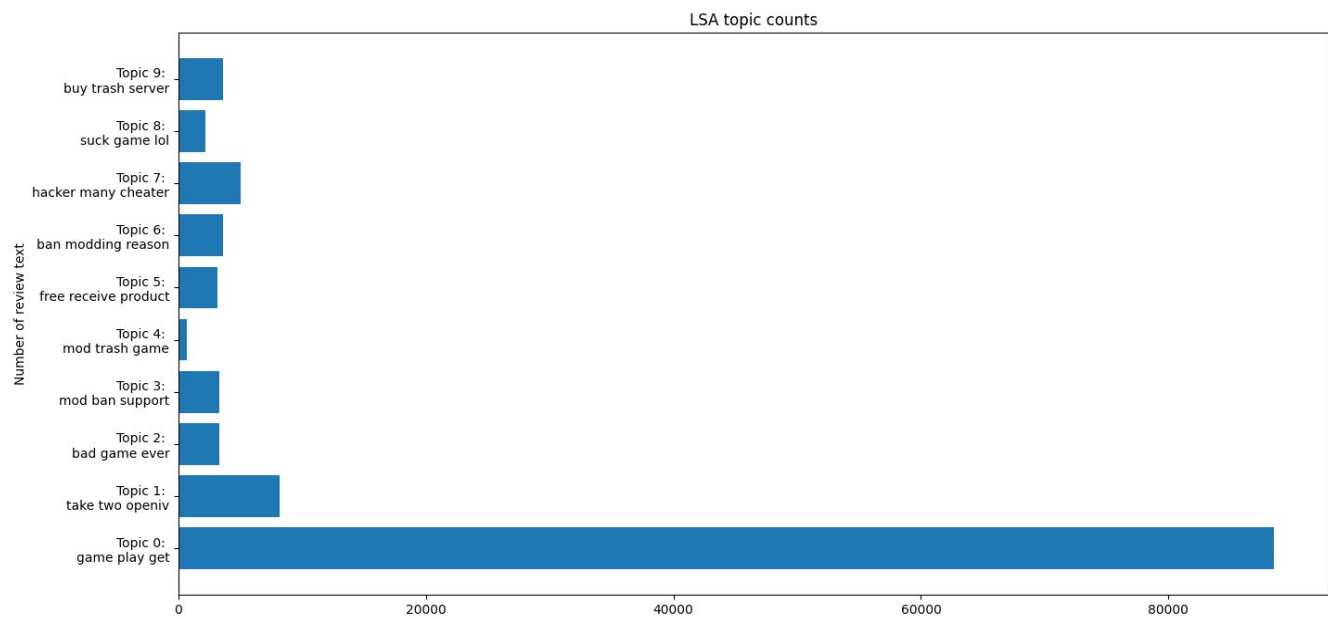
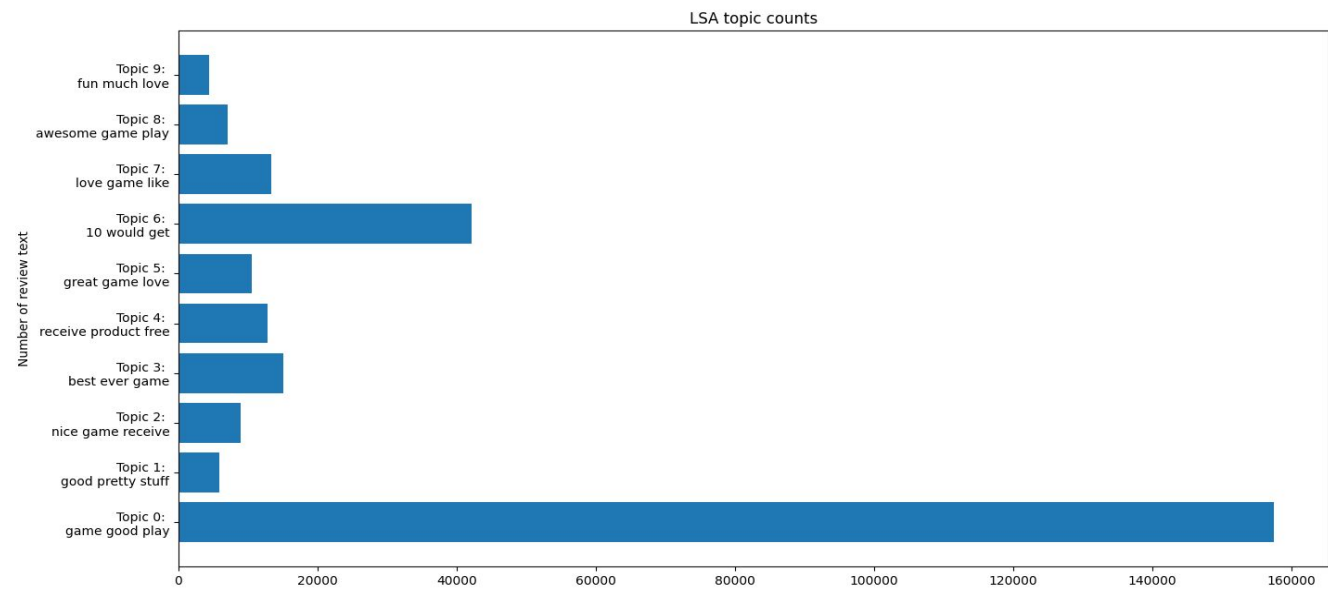


### trigramas mais comuns em análises recomendadas



### trigramas mais comuns em análises não recomendadas







# Problemas e Dificuldades

- Análise de Tópicos;
- Quantidade de Dados;
- Remoção de pontuações;
- Lematização.

## Próximos Passos

- Categorizar os jogos manualmente para análise entre similares;
- Melhorar procedimentos de limpeza e lematização de texto;
- Explorar outras opções de análises de tópicos, como o BERTopics.

# SteamDE: análises de *reviews* de jogos na Steam

# Obrigado!

Daniel de Viveiros Inácio  
Eduardo Darrazão



**STEAM®**



MINISTÉRIO DA  
**EDUCAÇÃO**

