

# Desafio Técnico

## A3Data

Felipe José Damascena Sousa





# O desafio: Situação Atual

**A Situação:** Uma editora parceira busca entender a performance de seus livros e autores através da análise de um grande volume de avaliações online.

**O Processo Atual:** A exploração desses dados é feita **manualmente** por uma equipe de **5 analistas**, que levam em média **3 dias** para concluir uma análise.

**O Custo:** Este método representa um custo salarial direto de **R\$5.000,00 por análise**, além de imobilizar uma equipe qualificada em tarefas repetitivas.

**A Dor Principal:** O processo é lento, caro e não escalável, dificultando a tomada de decisões ágeis e baseadas em dados.



# O desafio: Objetivo

**Objetivo Central:** Construir uma ferramenta que **automatize e otimize** o processo de análise de avaliações, transformando dados textuais em insights de negócio claros e acionáveis.

## **Metas Específicas:**

1. **Entender Performance:** Criar um método para avaliar, de forma rápida e precisa, a recepção de um autor, um gênero ou um livro específico.
2. **Identificar Influenciadores:** Desenvolver um sistema para encontrar usuários com opiniões detalhadas e relevantes, que possam ser contatados para entrevistas e pesquisas de mercado.
3. **Gerar Impacto Real:** Propor uma solução que **reduza drasticamente os custos e o tempo** da análise, permitindo que a equipe de especialistas foque em atividades estratégicas.



# Roadmap - Análise Exploratória e Hipóteses

**Objetivo:** Entender os dados, garantir sua qualidade e formular as perguntas estratégicas que guiarão o projeto.

Entrega	Detalhes	Duração
Dataset Unificado	Análise, limpeza e junção das bases de livros e reviews.	3 dias
Relatório de Hipóteses	<p>Geração do relatório de Análise Exploratória (EDA).</p> <p><b>H1:</b> Tem gêneros que em média recebem melhores médias que outras (ex: 'Ficção' e 'Biografia').</p> <p><b>H2:</b> Certos gêneros (ex: 'Ficção') têm notas mais polarizadas (5 ou 1) do que outros.</p> <p><b>H3:</b> O tamanho do texto da review tem correlação com a força da opinião (5 ou 1).</p>	1 dia



# Roadmap - Tópicos e Sentimentos

**Objetivo:** Desenvolver uma ferramenta funcional (POC) que automatiza a identificação dos principais temas e sentimentos das reviews, entregando insights acionáveis e imediatos.

Entrega	Detalhes	Métrica	Duração
Modelo de Tópicos	Extrair múltiplos temas de cada review utilizando BERTopic.	Relevância do Tópico (Similaridade Tópico-Review).  Avaliação humana.	3 dias
Modelo de Sentimento Treinado	Classificar o sentimento (positivo/negativo/neutro) para cada tema identificado usando modelo pré-treinado do HuggingFace.	F1 score. Podemos usar o rating para rotular ou rotular manualmente.	2 dias
Dashboard Interativo (POC)	Criar dashboard com os insights no Power BI	Feedback qualitativo dos analistas	2 dias



# Roadmap - LLM Chat

**Objetivo:** Transformar a análise de dados em um diálogo interativo. O propósito é criar uma interface de chat onde o analista pode "conversar" com as reviews, fazendo perguntas complexas ou pedindo por resumos em linguagem natural e recebendo respostas instantâneas, geradas por um LLM.

Entrega	Detalhes	Métrica	Duração
Chat de Análise (POC)	<p>Desenvolver uma interface onde o analista pode fazer perguntas abertas, como "Quais são as principais críticas ao final do livro?" ou "Me dê um resumo das opiniões sobre os personagens do autor".</p> <p>Retornar índices de reviews de exemplos quando necessário.</p> <p>Permitir na interface o analista escolher se quer conversar com reviews específicas de um gênero, autor, livro ou avaliação</p>	<p>Like/Dislike em relação à qualidade da resposta</p> <p>Feedback qualitativo dos analistas</p> <p>LLM como avaliador</p>	1 semana



# Roadmap - Produção e Monitoramento

**Objetivo:** Colocar as POCs validadas (testadas pelos stakeholders) em produção, garantindo que operem de forma estável, escalável e sempre atualizada, e medir continuamente seu impacto no negócio.

Entrega	Detalhes	Métrica	Duração
Pipeline de Dados em Tempo Real no Dashboard	Desenvolver uma arquitetura que captura novas reviews, processa e atualiza os dados automaticamente na base de dados e no dashboard.	Latência	2 dia
Monitoramento de Modelos	Implementar um sistema para detectar desvios (drift) e degradação do desempenho dos modelos em produção.	Sentimento: Rating + Humano (subset)  Tópicos: Similaridade.+ Humano (subset)  Chat: Usuário (Like/Dislike) + Feedback qualitativo dos usuários	1 semana
Medição de Impacto	Acompanhar a redução no tempo gasto pelos analistas e o aumento na geração de insights estratégicos.	Redução de horas em análise	-



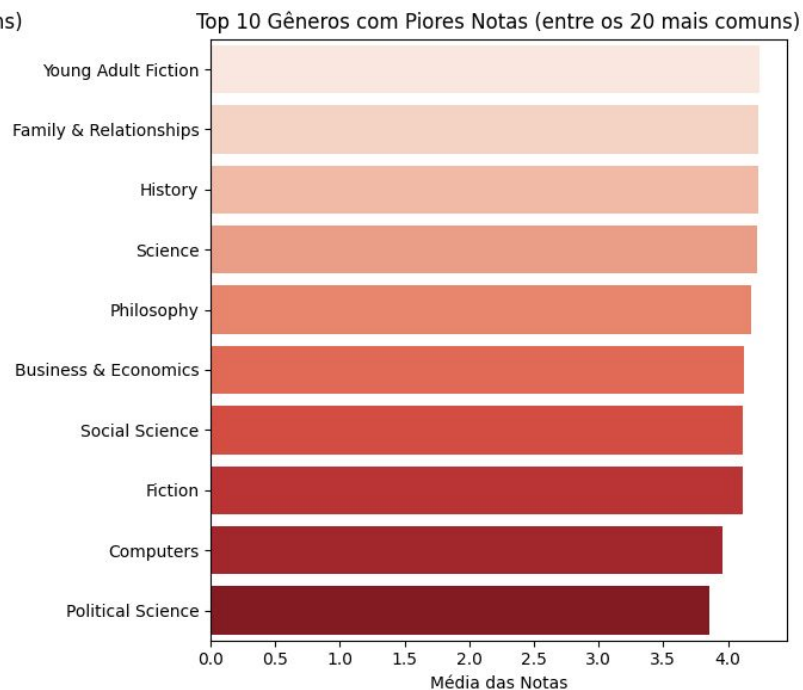
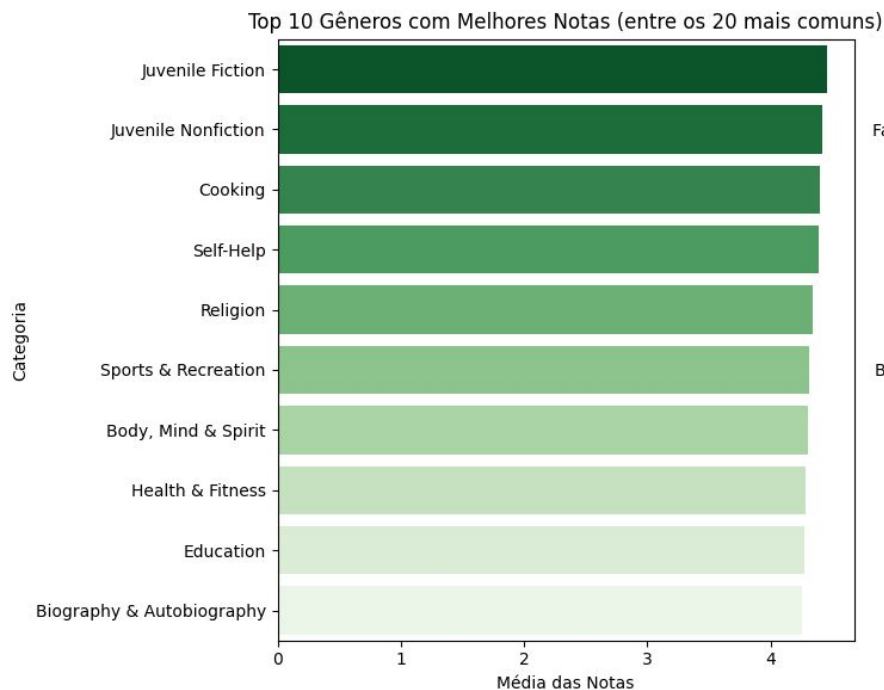
# Análise Exploratória

1. Temos o mesmo livro repetido algumas vezes no nos dados com título diferente.
  - a. Imaginei que a imagem poderia ter sido diferente, o que seria útil para analisar qual capa teve melhor vendas.
  - b. Como um dos objetivos da análise é avaliar a recepção de um livro, vou juntar os livros iguais após juntar o dataset atual com o de reviews. Vou usar a descrição, autor e data de publicação.
2. As datas de publicação tem alguns casos com apenas anos e outros com a data completa.
  - a. Se essas datas forem relevantes, eu iria tentar encontrar a data de publicação em alguma outra fonte de forma automática
  - b. Se não achar os dados, colocar tudo como ano caso não tenha essas, mas isso iria perder muita informação. Outra opção, dependendo da análise, usar os dados que tem mesmo.



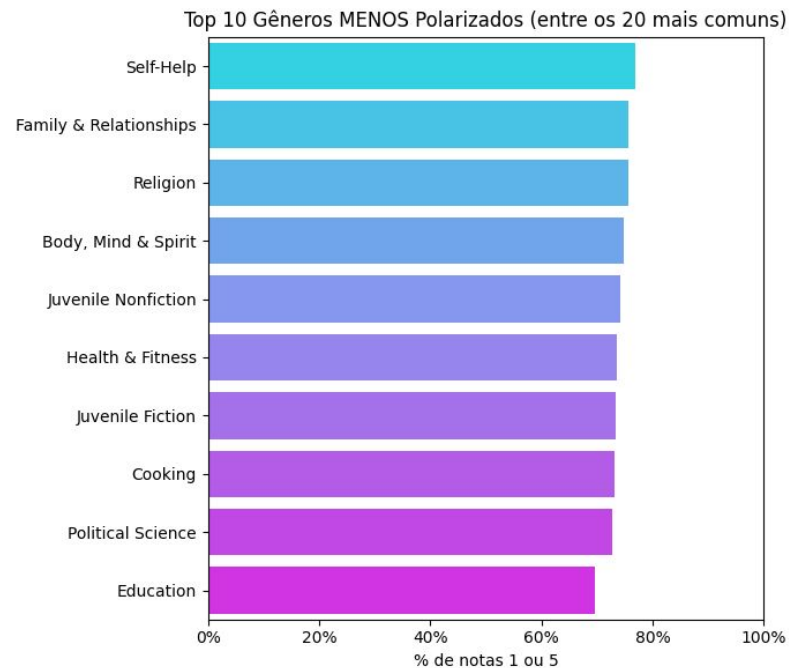
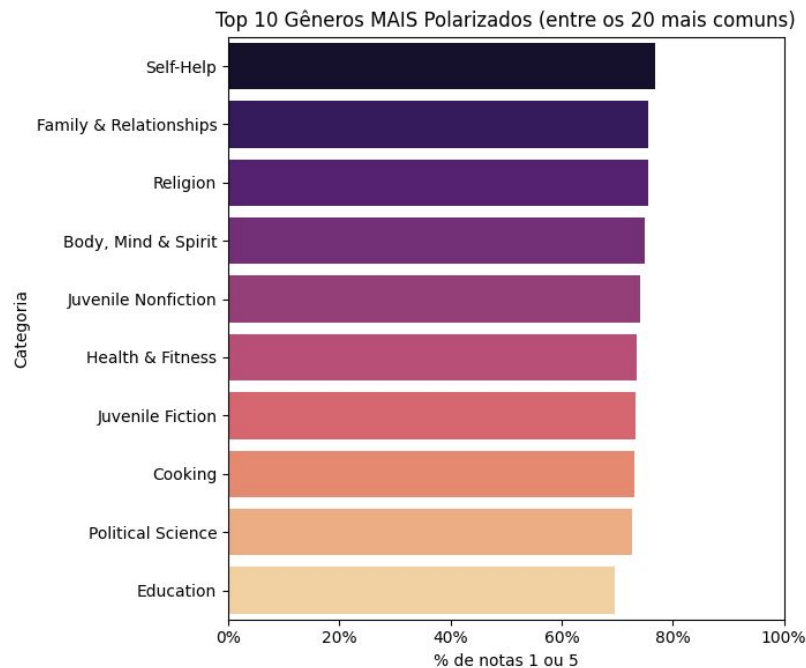


# Análise Exploratória



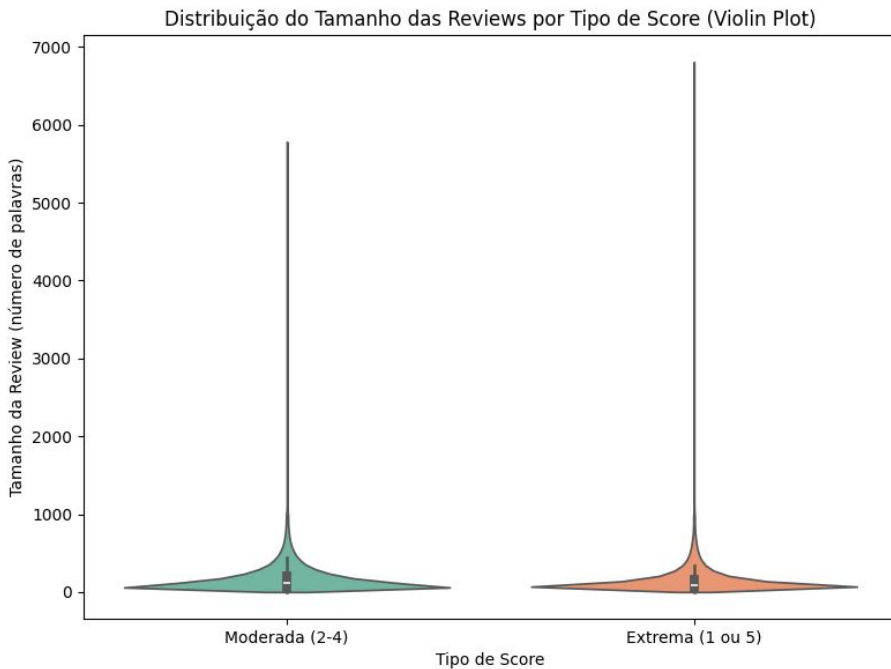


# Análise Exploratória





# Análise Exploratória



## Correlação:

Pearson: -0.0477

Spearman: -0.1075

Kendall : -0.0819

# Análise de Reviews

Dashboard Interativo de Tópicos e Sentimentos

## A Fúria dos Reis

por George R. R. Martin

Nota Média

4.4 ★

Total de Reviews

12,481

Sentimento Geral

82% Positivo

Gênero

Fantasia

### Principais Tópicos



Desenvolvimento de Personagens

4123



Intriga Política e Trama

3580



Ritmo do Livro

2950



### Análise Detalhada: Desenvolvimento de Personagens

Filtrando 4123 menções

Todos

Positivo

Negativo

LeitorVoraz22

Nota: 5 ★

A evolução do Tyrion neste livro é simplesmente genial. Martin constrói seus personagens com uma profundidade psicológica incrível, cheios de falhas e virtudes.

Ana\_Leitora

Nota: 5 ★





# Conclusão

## O Dashboard de Tópicos e Sentimentos (POC 1)

- **O que ele resolve?** A necessidade de entender o "quadro geral" (o *big picture*).
- **Como?** Agrega visualmente milhares de reviews, mostrando os temas mais discutidos (ex: "Desenvolvimento de Personagens") e a distribuição de sentimentos para cada um.
- **Impacto Direto: Elimina a leitura e tabulação manual.** Em vez de passar dias lendo, o analista tem uma visão completa e priorizada em segundos, respondendo à pergunta: *"Sobre o que nossos leitores estão falando e como se sentem sobre isso?"*.

## O Chatbot Conversacional com LLM (POC 2)

- **O que ele resolve?** A necessidade de fazer análises específicas e encontrar usuários.
- **Como?** Permite ao analista "conversar" com os dados, fazendo perguntas complexas em linguagem natural, como: *"Encontre leitores que gostaram da trama, mas acharam o final fraco"*.
- **Impacto Direto: Elimina a busca manual por agulha no palheiro.** O tempo para encontrar perfis específicos para entrevistas ou entender um nicho de opinião cai de horas para segundos.