



## Teste Técnico CD – NLP/LLM

O teste técnico para a vaga de Cientista de Dados (LLM/NLP) da A3Data consiste na exploração de uma base de dados com avaliações de livros e alguns metadados sobre os livros, acesse os dados [aqui](#) e [aqui](#).

O objetivo deste teste é que você consiga explorar e demonstrar suas habilidades técnicas e de negócio, trazendo insights acionáveis para o negócio. **Para esse projeto, o conhecimento em NLP é essencial e a utilização de LLMs é um diferencial. Nesta base temos colunas textuais e esperamos que você faça bom uso delas mostrando suas habilidades.**

### **Definição do problema:**

A editora contratou a A3Data para explorar uma base de avaliações de livros. Atualmente o processo de exploração é feito manualmente pelos analistas e leva muito tempo. O objetivo é tanto de entender a performance de um certo autor, um gênero ou até buscar usuários com opiniões relevantes para uma entrevista. A editora gostaria de uma ferramenta que agilizasse e otimizasse este processo. Você pode desenvolver o teste em sua linguagem de programação de preferência.

Considere que atualmente o time leva cerca de 3 dias para fazer uma análise e possui 5 pessoas para fazer esse trabalho manualmente, com um custo salarial de R\$5000,00. Tente apresentar como o modelo faria um impacto na redução desse custo, permitindo que os analistas fossem alocados em outras áreas diferentes.

**Você deverá entregar os seguintes artefatos:**

1. Apresentação dos resultados, um Power Point ou Google slides, simulando uma apresentação para o time de negócio. Ela deve conter:
  - a. Apresentação do desafio
  - b. Planejamento de entregáveis (inclusive futuros) – Roadmap
  - c. Explicação do processo utilizado
  - d. Hipóteses levantadas
  - e. Análise exploratória.
  - f. Sumarização das informações textuais (Pelo menos a proposta com utilização de alguma tecnologia).
  - g. Uso de bases de conhecimento (Proposta de utilização).
  - h. Métricas para avaliar a qualidade do resultado.

- i. Estimativa de impacto das ações sugeridas (financeiras ou de melhoria em processos).
  - j. Desejável/Opcional: POC da solução
2. Código no GitHub

**O que iremos avaliar:**

- 1. Capacidade analítica
- 2. Qualidade do storytelling na apresentação do problema e no código
- 3. Qualidade e clareza do código
- 4. Criatividade
- 5. Planejamento e organização
- 6. Capacidade de se conectar com o negócio
- 7. Ser propositivo

Você irá apresentar o desenvolvimento deste desafio para a equipe da A3Data. Esta apresentação será realizada durante uma entrevista em um tempo máximo de 45 minutos (30 minutos apresentação + 15 para perguntas).