





# Software Analytics Canvas

Projeto: Conversão de Linguagem Natural em Operações de Análise de Dados com apoio de Modelos de Inteligência Artificial Generativa

Autor: Frederico Comério

<div> <b>1. Question</b></div> <div><i>What is it that we want to know about the software / processes / usage / organization / etc.?</i></div> <div>Conversão de Linguagem Natural em Operações de Análise de Dados com Apoio de Modelos de Inteligência Artificial Generativa.</div> <div>O objetivo principal é avaliar como a integração de um modelo de linguagem natural (LLM), como o GPT-4, pode facilitar a análise de dados para usuários sem conhecimento técnico, permitindo que eles interajam com datasets complexos através de uma interface de chat.</div>	<div> <b>2. Data Sources</b></div> <div><i>Which data can possibly answer our question? What information do we need?</i></div> <div>IMDB 5000 Movie Dataset (disponível no Kaggle), que serve como base para as análises. Logs de interação dos usuários com a interface para entender as perguntas mais frequentes e a efetividade das respostas geradas pelo modelo. Análises dos resultados retornados pelo modelo GPT-4 para verificar a precisão na execução dos comandos solicitados em linguagem natural.</div>	<div> <b>3. Heuristics</b></div> <div><i>Which assumptions do we want to make to simplify the answer to our question?</i></div> <div>Algumas das suposições incluem:  O modelo de linguagem natural consegue compreender corretamente a intenção dos usuários e gerar código Pandas adequado para as operações solicitadas. Os usuários conseguem formular perguntas em linguagem natural com um mínimo de clareza e contexto para que o modelo possa interpretar e gerar resultados corretos.</div>	<div> <b>4. Validation</b></div> <div><i>What results do we expect from our analysis, how are they reviewed and presented in an understandable way?</i></div> <div>Espera-se que o modelo execute corretamente as operações solicitadas em linguagem natural, com alta precisão. Para validação, utilizamos um caderno de testes com 10 perguntas que exigem diferentes níveis de complexidade nas operações de análise de dados. Os resultados são revisados comparando as respostas geradas pelo modelo com as respostas esperadas. Um score de 100% de precisão foi obtido nos testes realizados.</div>
---	---	---	---



## 5. Implementation

*How can we implement the analysis step by step and in a comprehensible way?*

Integração do GPT-4 com a biblioteca PandasAI para converter perguntas em linguagem natural em código Pandas.

Desenvolvimento de uma interface web usando Streamlit para que os usuários possam carregar datasets e interagir através de um chat.

Implementação de mecanismos para coleta de logs e feedback dos usuários para ajustes contínuos no sistema.

Realização de testes e ajustes nas heurísticas de prompts para melhorar a precisão das respostas.



## 6. Results

*What are the main insights from our analysis?*

Os principais insights são:

O modelo GPT-4 foi capaz de interpretar corretamente as perguntas e realizar operações complexas com alta precisão. A interface desenvolvida democratiza o acesso à análise de dados para usuários não técnicos, permitindo que eles explorem datasets complexos sem precisar aprender programação.



## 7. Next Steps

*What follow-up actions can we derive from the findings? Who or what do we need to address next?*

Explorar a integração com outras bibliotecas de visualização de dados, como Seaborn e Plotly, para enriquecer a experiência do usuário. Implementar melhorias no tratamento de linguagem ambígua e gramaticalmente incorreta. Pesquisar mecanismos de feedback contínuo do usuário para ajustar os modelos e os prompts automaticamente. Avaliar a escalabilidade da solução para diferentes domínios além da análise de dados, como BI e automação de processos.