

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA

ELÉTRICA E DE COMPUTAÇÃO

Consultando Base de Dados Com Linguagem Natural

Disciplina: Tópicos Especiais em Machine Learning

Professor: Ivanovitch Medeiros Dantas da Silva

Alunos: Hagi Jakobson Dantas da Costa

Mikael Pereira Belo Themoteo

Maio de 2023

Descrição do Problema

Busca e seleção de dados de valor

- Uma das responsabilidades de um Analista/Cientista de dados é extrair informações valiosas armazenados nas bases de dados, fornecendo insights estratégicos para impulsionar as decisões de negócios.



Descrição do Problema

Busca e seleção de dados de valor

- Para desempenhar este papel, evitar gargalos e garantir uma resposta eficiente às demandas de negócios, é importante gerenciar:
 - As requisições;
 - A complexidade das consultas;
 - E a disponibilidade dos analistas de dados.

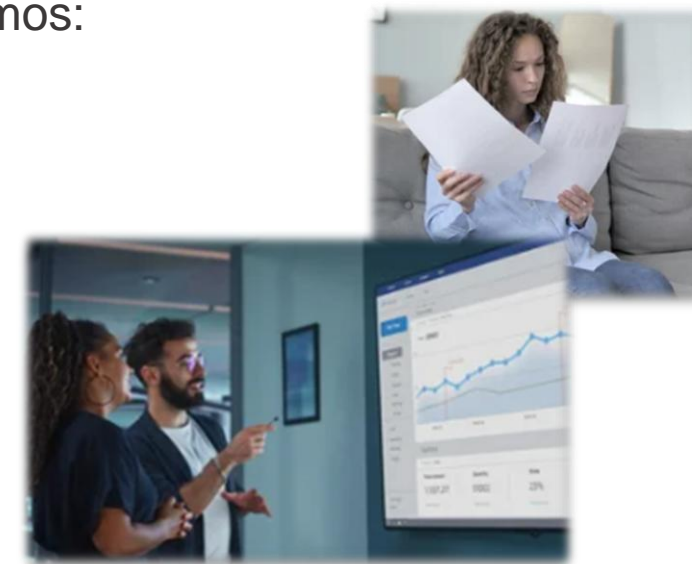


Solução Proposta

Busca de informação através de linguagem natural

- Uma das primeiras formas de vislumbrar e propor a solução para um problema é descrevendo-o e como imaginamos os passos até a resolução final. De forma intuitiva, podemos:

- Descrever verbalmente, ou anotar em tópicos, os passos mais evidentes;
- Apontar as fontes de dados disponíveis;
- Selecionar o conjunto de dados que nos ajude a solucionar o problema;
- Gerar informação e insights a partir das análises.



Solução Proposta

Geração de consultas SQL utilizando linguagem natural

- Através da implementação de uma solução (Web App) que utiliza Large Language Models (LLMs) em conjunto com o framework LangChain. Essa abordagem permite que usuários, mesmo sem conhecimentos de SQL, sejam capazes de fazer consultas a bases de dados tabulares utilizando linguagem natural.



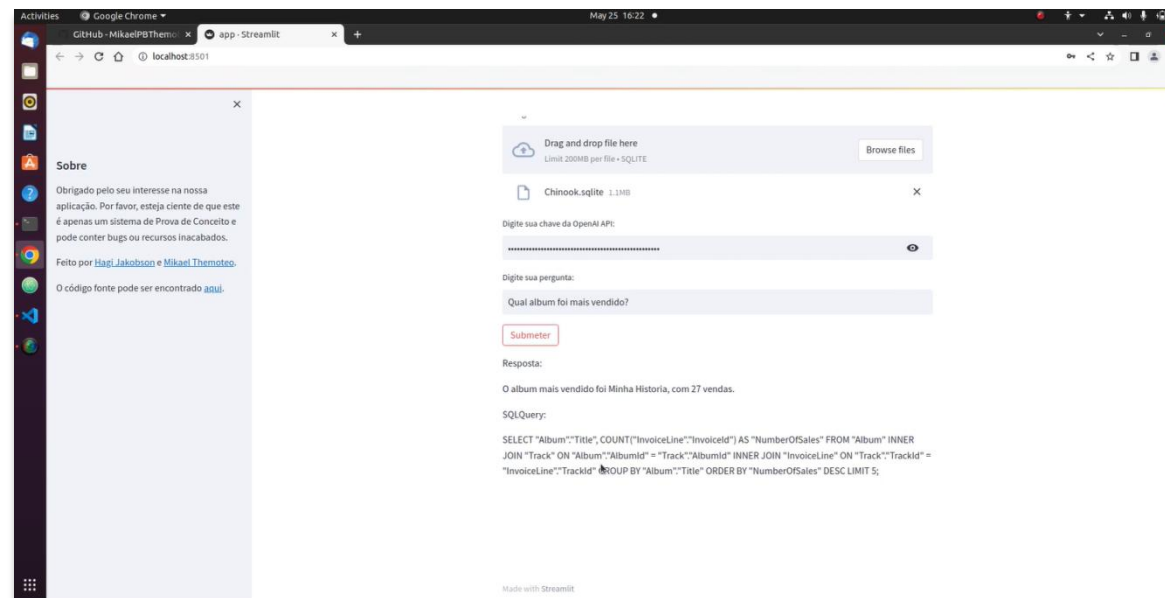
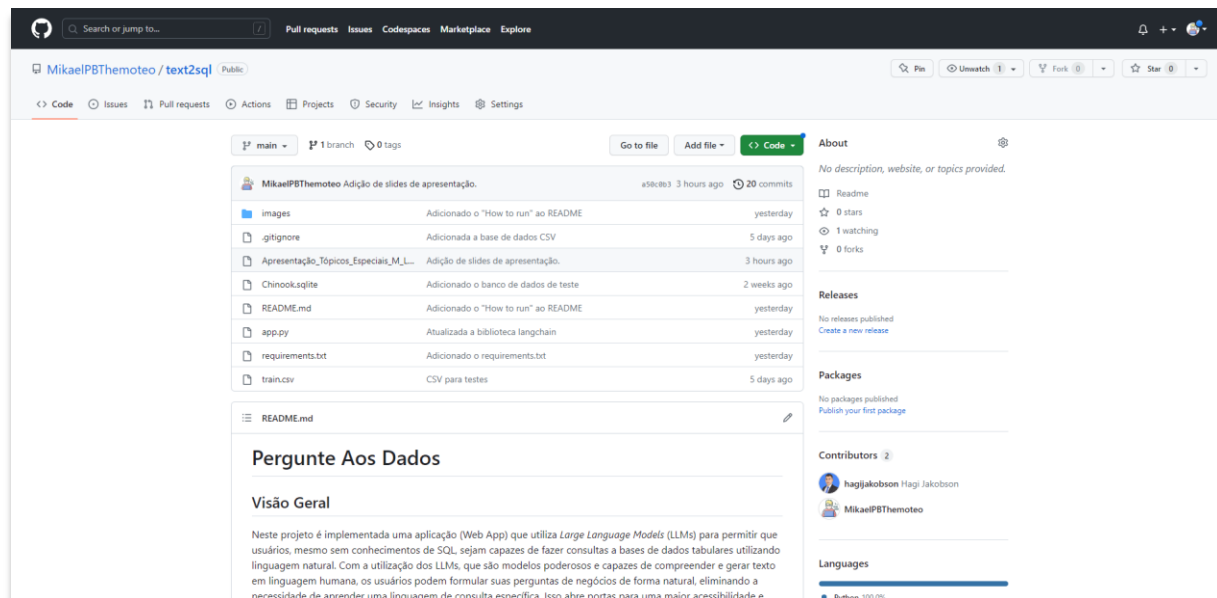
Solução Proposta

Geração de consultas SQL utilizando linguagem natural

- O aplicativo foi desenvolvido com as bibliotecas LangChain e Streamlit e a API da OpenAI. O objetivo é permitir que os usuários façam perguntas sobre suas bases de dados usando linguagem natural e obtenham respostas correspondentes.

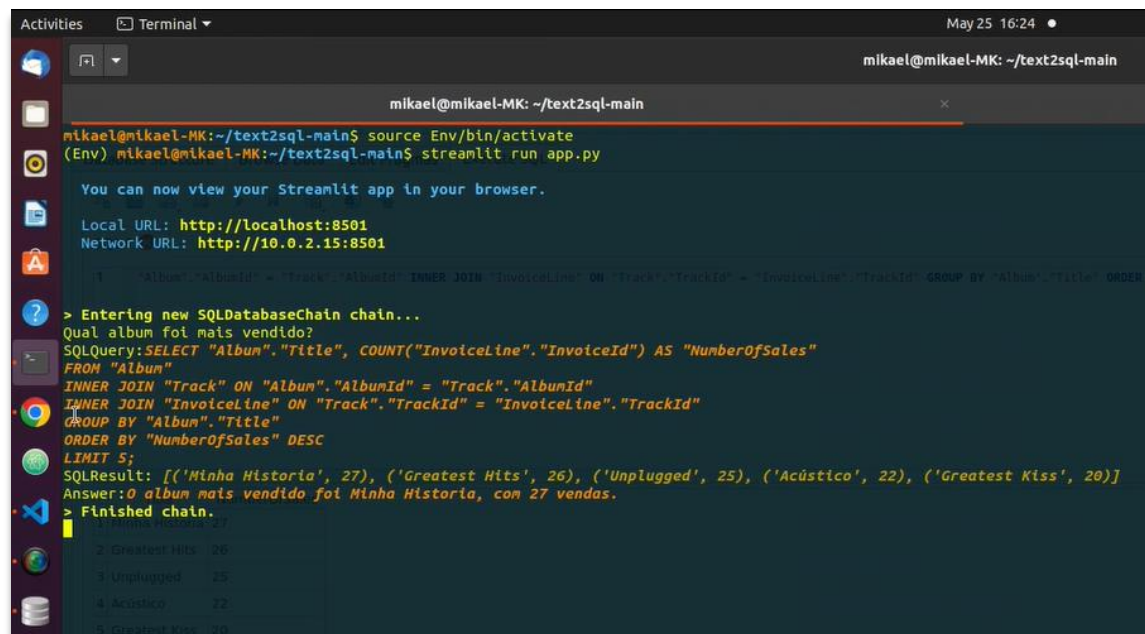


Solução Proposta



<https://github.com/MikaelPBThemoteo/text2sql>

Solução Proposta



Activities Terminal May 25 16:24

mikael@mikael-MK: ~/text2sql-main

```
mikael@mikael-MK:~/text2sql-main$ source Env/bin/activate
(Env) mikael@mikael-MK:~/text2sql-main$ streamlit run app.py

You can now view your Streamlit app in your browser.

Local URL: http://localhost:8501
Network URL: http://10.0.2.15:8501

1 "Album"."AlbumId" = "Track"."AlbumId" INNER JOIN "InvoiceLine" ON "Track"."TrackId" = "InvoiceLine"."TrackId" GROUP BY "Album"."Title" ORDER
> Entering new SQLiteDatabaseChain chain...
Qual album foi mais vendido?
SQLQuery:SELECT "Album"."Title", COUNT("InvoiceLine"."InvoiceId") AS "NumberOfSales"
FROM "Album"
INNER JOIN "Track" ON "Album"."AlbumId" = "Track"."AlbumId"
INNER JOIN "InvoiceLine" ON "Track"."TrackId" = "InvoiceLine"."TrackId"
GROUP BY "Album"."Title"
ORDER BY "NumberOfSales" DESC
LIMIT 5;
SQLResult: [(('Minha Historia', 27), ('Greatest Hits', 26), ('Unplugged', 25), ('Acústico', 22), ('Greatest Kiss', 20))]
Answer:0 album mais vendido foi Minha Historia, com 27 vendas.
> Finished chain.
```

	Title	NumberOfSales
1	Minha Historia	27
2	Greatest Hits	26
3	Unplugged	25
4	Acústico	22
5	Greatest Kiss	20

