

NSYNC

Sprint 3

Nossa Equipe



Ana Goes



Gabriel Coletto



Gabriel Farias



Hugo Noyma



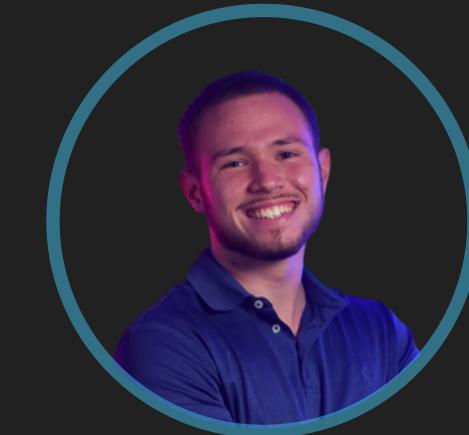
João Paulo



Lucas Nogueira



Mauro Chagas



Vitto Mazeto

Tópicos de hoje:

- Preparação dos Dados
- Transcrição dos Vídeos
- Melhorias no Modelo (Fine tuning)
- Contexto para o Chatbot (RAG)
- Analise Semântica
- Evolução do Artigo

Preparação dos Dados



Problema

451 linhas (30 colunas)
118 SKUs, mas 73% duplicatas
(tamanhos separados)
Tipos de dados inconsistentes



O que fizemos

Padronização de tipos
Consolidação por SKU
Novas colunas

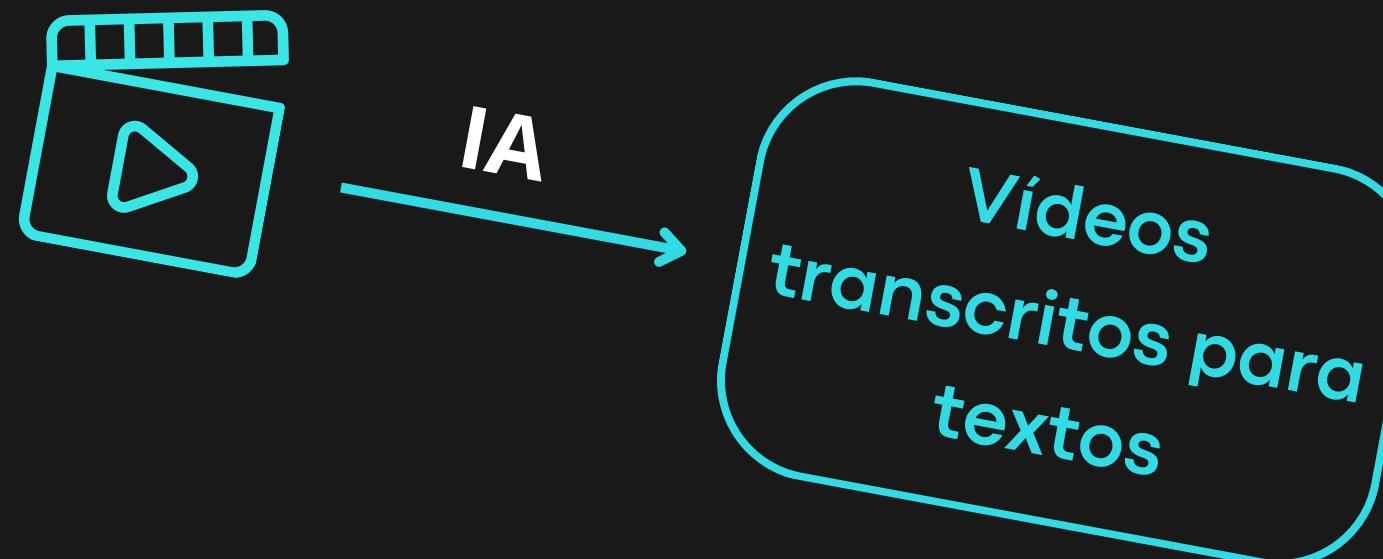


Resultado

118 registros únicos (de
451)
73% de redundância,
100% da info preservada

Base pronta para RAG: consultas mais precisas, sem repetição de produtos.

Transcrição dos Vídeos



Personalização real: captura
vocabulário e tom da Bia

Cobertura total: funciona mesmo
sem legendas

Consistência: qualidade uniforme
nas transcrições

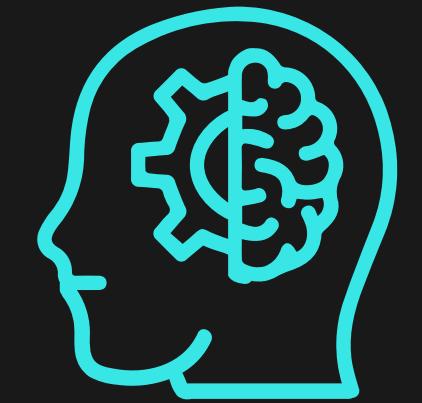


18 vídeos processados
55.5k caracteres
consolidados
Precisão alta ($\approx >95\%$)

Fine-Tuning



Ideia: Ensinar o modelo a “falar
Curadobia”



É como treinar um estilista que já
conhece moda, mas precisa aprender o
jeito de falar da Bia.

Resultados do Fine-Tuning



O modelo aprendeu o vocabulário e tom da Bia

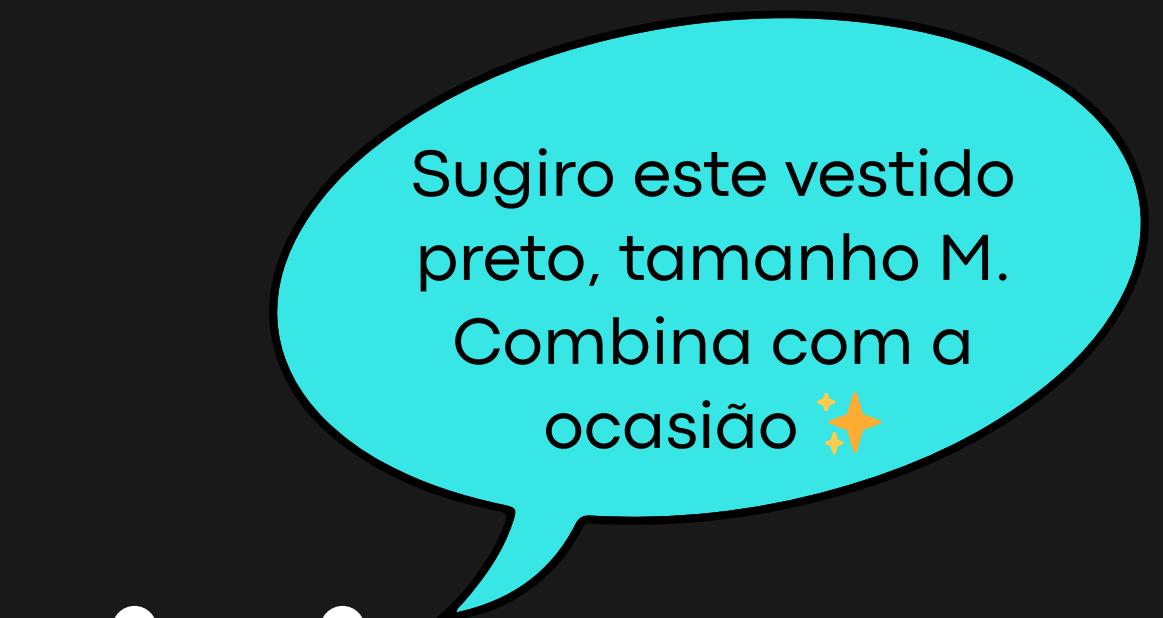


Bom em perguntas frequentes de moda



Ainda precisa melhorar em casos raros ou complexos

RAG



Análise Semântica

Ideia: Entender o sentido por trás das perguntas.

Como organizar um guarda-roupa — peças diferentes, mas que combinam em estilo.



Análise Semântica

"Qual tamanho tem essa rupa?"

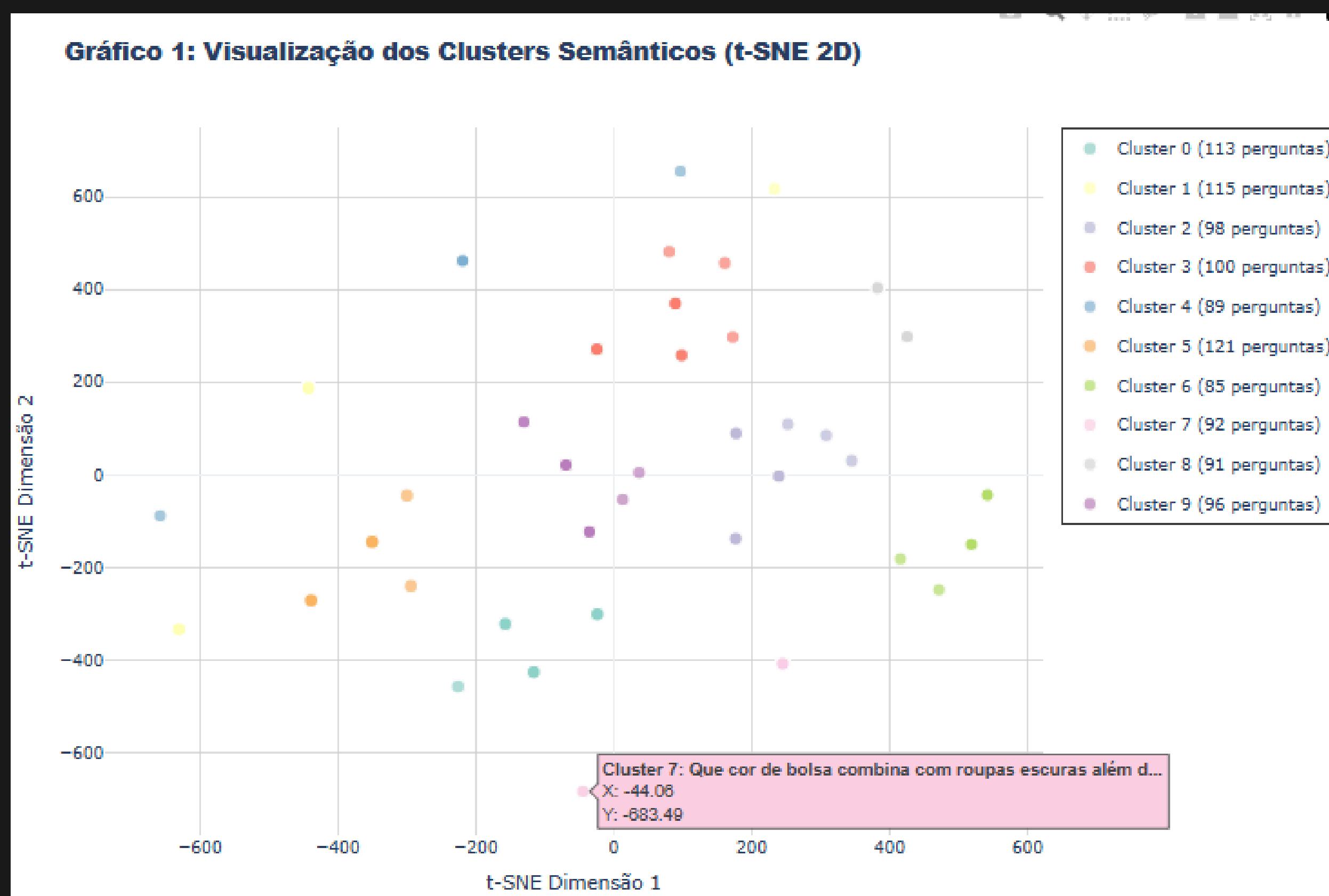
Grupo de dúvida sobre

"Essa calça é P de modelagem grande?"

tamanho

"Quero trocar essa roupa"

Grupo sobre trocas

Gráfico 1: Visualização dos Clusters Semânticos (t-SNE 2D)

Evolução do Artigo



Materiais e Métodos



Resultados



Análise e Discussão



Obrigado!