

Meu Plano de Ação: Como Vou Fazer os 3 Experimentos

Este é o meu plano de como vou executar os três experimentos que escolhi, explicado do meu jeito, passo a passo.

Experimento 1: Palavras e Temas Dominantes

Meu objetivo: Descobrir quais palavras a banda mais usa e o que isso revela. Vou usar três métodos: contagem simples, TF-IDF e uma nuvem de palavras.

Como eu vou fazer:

1. **Primeiro, vou preparar os dados:** Antes de contar, vou "limpar" todas as letras. Isso significa que vou remover palavras muito comuns que não dizem nada (como "e", "de", "para", "que", "o", "a", "um"), e também vou deixar tudo em letra minúscula.
2. **Depois, vou fazer a Contagem Simples:** Vou contar quantas vezes cada palavra (já limpa) aparece em *todas* as músicas. Vou pegar as 10 ou 20 mais frequentes. Isso vai me dar o vocabulário geral da banda.
3. **Em seguida, vou usar o TF-IDF:** Esse é um pouco diferente. Ele não vai me dizer só a palavra mais frequente, mas a palavra que é *mais importante* para uma *música específica*. Por exemplo, se "pampa" aparece muito em uma música, mas quase nunca nas outras, o TF-IDF vai me mostrar que "pampa" é um termo super relevante *para aquela* música. Vou tentar listar os top 5 termos TF-IDF para algumas músicas.
4. **Por fim, a Nuvem de Palavras:** Vou usar o resultado da minha "Contagem Simples" (do passo 2) para gerar uma imagem visual. Nela, as palavras que aparecem mais vezes vão ficar maiores, e as que aparecem menos, menores. É uma forma rápida de "bater o olho" e ver os temas principais.
5. **Análise:** Vou olhar para essas listas e para a nuvem e pensar: "O que essas palavras dizem sobre a banda? São palavras mais filosóficas, mais políticas, mais românticas?"

Experimento 2: Classificação Intuitiva

Meu objetivo: Tentar classificar as letras como "melancólicas", "otimistas" ou "filosóficas". Vou fazer isso usando regras bem simples que eu mesmo vou criar.

Como eu vou fazer:

1. **Primeiro, vou criar meus "dicionários":** Vou criar três listas de palavras-chave para cada sentimento:
 - **Lista Melancólica:** Palavras como "tristeza", "dor", "sozinho", "fim", "solidão", "chuva", "noite".
 - **Lista Otimista:** Palavras como "sol", "esperança", "sonho", "luz", "amanhã", "sorriso",

"novo".

- Lista Filosófica: Palavras como "vida", "tempo", "razão", "ser", "mundo", "limite", "porquê".

(Vou adicionar mais palavras que eu lembrar).

2. **Depois, vou processar cada letra:** Vou mandar meu script ler cada música.
3. **Em seguida, vou contar as palavras de cada lista:** Para cada música, vou verificar e anotar:
 - Quantas palavras da minha Lista Melancólica ela tem? (Ex: contagem_melancolica = 5)
 - Quantas da Lista Otimista? (Ex: contagem_otimista = 2)
 - Quantas da Lista Filosófica? (Ex: contagem_filosofica = 8)
4. **Aí, vou decidir a classificação:** Minha regra será: a categoria que tiver a **maior contagem** vence. No meu exemplo acima, a música seria classificada como **"Filosófica"**.
5. **Análise:** Quando todas as músicas estiverem classificadas, vou olhar a lista. As músicas que eu conheço bem (como "Pra Ser Sincero" ou "Infinita Highway") caíram na categoria que eu esperava? Vou ver onde esse método acertou e onde errou.

Experimento 3: Similaridade entre Letras

Meu objetivo: Encontrar músicas que falam de temas parecidos, mesmo que elas não usem exatamente as mesmas palavras (por exemplo, "solidão" e "isolamento").

Como eu vou fazer:

1. **Primeiro, vou escolher a ferramenta (modelo):** Vou usar um modelo de "embedding" já pronto, provavelmente da biblioteca sentence-transformers. Vou pegar um que entenda bem português.
2. **Depois, vou "traduzir" as letras em números:** Vou explicar de forma simples: esse modelo vai "ler" cada letra e me devolver uma lista de números (um "vetor"). A mágica é que letras com significados parecidos vão ter listas de números (vetores) parecidas.
3. **Em seguida, vou gerar os vetores:** Vou passar *todas* as letras do meu dataset por esse modelo. No final, eu terei um "vetor" guardado para cada música.
4. **Agora, vou calcular a similaridade:** Vou usar uma fórmula matemática (tipo "similaridade de cosseno") que compara duas listas de números e me diz o quão "próximas" elas estão (normalmente um número de 0 a 1).
5. **Por fim, vou testar:** Essa é a parte mais legal. Vou fazer perguntas ao meu programa:
 - "Pegue a música 'Infinita Highway'. Agora, me mostre as 5 músicas do dataset que são mais parecidas com ela (que têm a maior pontuação de similaridade)."
 - Aí eu vou ler o nome dessas 5 músicas. Elas realmente falam de temas parecidos (estrada, liberdade, tempo)? Vou repetir isso para várias músicas e descobrir os "grupos" de temas que a máquina encontrou.