

# Agrupamento e Redução de Dimensionalidade

[Iniciar tarefa](#)

**Vencimento** 14 nov por 23:59      **Pontos** 100  
**Enviando** uma caixa de entrada de texto ou um URL de site  
**Disponível** 7 nov em 0:00 - 16 nov em 23:59

## Descrição do problema:

Nesse laboratório iremos aplicar os conceitos de agrupamento e redução de dimensionalidade em dados textuais, mais especificamente, em dados de letras de música. A ideia é encontrar grupos em letras de músicas de três gêneros musicais.

## Link para os dados:

Os dados para este laboratório foram coletados pelo professor Nazareno Andrade, estando disponíveis no [GitHub](https://github.com/nazareno/palavras-nas-letras) [↗](https://github.com/nazareno/palavras-nas-letras) [\\_](https://github.com/nazareno/palavras-nas-letras) [. São eles:](https://github.com/nazareno/palavras-nas-letras)

- Rock: [link](https://raw.githubusercontent.com/nazareno/palavras-nas-letras/master/letras-ptbr-rock-grande.csv) [↗](https://raw.githubusercontent.com/nazareno/palavras-nas-letras/master/letras-ptbr-rock-grande.csv) [\\_](https://raw.githubusercontent.com/nazareno/palavras-nas-letras/master/letras-ptbr-rock-grande.csv) [. São eles:](https://raw.githubusercontent.com/nazareno/palavras-nas-letras/master/letras-ptbr-rock-grande.csv)
- Samba: [link](https://raw.githubusercontent.com/nazareno/palavras-nas-letras/master/letras-ptbr-samba-grande.csv) [↗](https://raw.githubusercontent.com/nazareno/palavras-nas-letras/master/letras-ptbr-samba-grande.csv) [\\_](https://raw.githubusercontent.com/nazareno/palavras-nas-letras/master/letras-ptbr-samba-grande.csv) [. São eles:](https://raw.githubusercontent.com/nazareno/palavras-nas-letras/master/letras-ptbr-samba-grande.csv)
- Sertanejo: [link](https://raw.githubusercontent.com/nazareno/palavras-nas-letras/master/letras-ptbr-sertanejo-grande.csv) [↗](https://raw.githubusercontent.com/nazareno/palavras-nas-letras/master/letras-ptbr-sertanejo-grande.csv) [\\_](https://raw.githubusercontent.com/nazareno/palavras-nas-letras/master/letras-ptbr-sertanejo-grande.csv) [. São eles:](https://raw.githubusercontent.com/nazareno/palavras-nas-letras/master/letras-ptbr-sertanejo-grande.csv)

## Descrição dos dados:

- **SName:** Coluna que contém o título da música
- **Lyric:** Coluna que contém a letra da música
- **Artist:** Nome do artista ou banda
- **Songs:** Número de músicas do artista. Caso o mesmo artista tenha mais de uma música no conjunto, o valor dessa coluna estará repetido em todas as músicas.
- **Popularity:** Popularidade do artista. Caso o mesmo artista tenha mais de uma música no conjunto, o valor dessa coluna estará repetido em todas as músicas.
- **Genre:** Gênero da música.
- **Genres:** Gêneros possíveis.

## As atividades esperadas para essa etapa são descritas a seguir:




Escolha um ou mais conjuntos de letras de músicas e realize as etapas descritas a seguir:

1. Utilize o k-means para procurar grupos de músicas, identificando o melhor número de grupos através das técnicas explicadas nas aulas. Em seguida, descreva os grupos encontrados

atraves das técnicas explicadas nas aulas. Em seguida, descreva os grupos encontrados, listando as palavras e os artistas mais frequentes em cada grupo. Tente nomear os grupos e explique em detalhes o racional para sua nomeação para cada grupo. Repita este processo utilizando como entrada para o k-means:

1. O título da música;
  2. A letra da música;
  3. A concatenação do título e letra;
2. Mesmo que 1 acima, mas aplique uma redução de dimensionalidade utilizando o PCA, e utilize os dados transformados como entrada para o k-means.
3. Compare a qualidade dos grupos com e sem PCA usando o coeficiente de silhueta.
4. Utilize o PCA e t-SNE e gere visualizações dos grupos encontrados nas questões anteriores. As visualizações indicam que os grupos são bem separados? Por que você acha que sim ou que não?

**Os seguintes vídeos e notebooks do Prof. Nazareno sobre agrupamento em dados de texto podem ajudar:**

- [Agrupamento de letras de Samba - Parte 1](https://youtu.be/dqNirYNBOYo)  [\\_ \(https://youtu.be/dqNirYNBOYo\)](https://youtu.be/dqNirYNBOYo)
- [Agrupamento de letras de Samba - Parte 2](https://youtu.be/YxmrLrpNaPY)  [\\_ \(https://youtu.be/YxmrLrpNaPY\)](https://youtu.be/YxmrLrpNaPY)
- [Notebook para os vídeos acima](https://colab.research.google.com/drive/1htf0_ZjnwM84NCrkeyBM6qhE_b2cbzZ0#revisionId=0Bxo3d)  [\\_ \(https://colab.research.google.com/drive/1htf0\\_ZjnwM84NCrkeyBM6qhE\\_b2cbzZ0#revisionId=0Bxo3d\)](https://colab.research.google.com/drive/1htf0_ZjnwM84NCrkeyBM6qhE_b2cbzZ0#revisionId=0Bxo3d)

**Como devo entregar?**

Link para o seu notebook no Colab.

