Aprendemos nesta aula a interpretar e construir as principais visualizações voltadas para distribuição dos dados. Entretanto, existe uma série de gráficos que podemos utilizar para este fim enriquecendo o nosso *storytelling*. Uma das formas é por meio do chamado **word cloud**, ou nuvem de palavras, que representa a distribuição de dados categóricos por meio das palavras-chave.

Através das nuvens de palavras, é possível ter uma visão rápida e intuitiva das palavras mais relevantes e sua frequência em um conjunto de dados representados pelos tamanhos que elas são exibidas na imagem.

Para criar esta visualização utilizando a linguagem Python, precisamos recorrer a biblioteca WordCloud, que já vem instalada de maneira nativa no Google Colab. Para testarmos a biblioteca, vamos recorrer ao projeto que fizemos nas primeiras aulas das vendas das lojas de departamento e desenvolver a nossa nuvem de palavras com os estados que mais aparecem no dataset.

### **Nuvem de palavras dos estados**

Primeiro importamos as bibliotecas que vamos utilizar. Um destaque para o método WordCloud() da biblioteca wordcloud responsável por gerar a nuvem de palavras e a função Image da biblioteca PIL(pillow) para podermos utilizar imagens dentro na nuvem de palavras (no exemplo a seguir).

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from wordcloud import WordCloud

from PIL import Image

Agora, vamos importar nossos dados, passando para um dicionário, como chaves, os nomes dos estados e, como valores, a contagem de repetições de cada estado na coluna estado. Na sequência criamos o objeto WordCloud() e geramos uma nuvem de palavras com o dicionário que criamos, passando-o como parâmetro do sub-método generate\_from\_frequencies:

# Importando o relatório de vendas

vendas = pd.read\_csv("https://raw.githubusercontent.com/alura-cursos/dataviz-graficos/master/dados/relatorio\_vendas.csv")

# Gerando um dicionário de frequências

dic = dict(vendas["estado"].value\_counts())

# Criar um objeto wordcloud

wordcloud = WordCloud()

# Cria uma wordcloud através do dicionário de frequências

wordcloud.generate\_from\_frequencies(frequencies = dic)

fig, ax = plt.subplots(figsize = (15, 10))

plt.imshow(wordcloud, interpolation = 'bilinear')

ax.axis("off") # retirando bordas e ticks

plt.show()

Por fim, definimos o tamanho da figura que vamos desenhar e exibimos a nuvem de palavras, retirando as bordas e elementos gráficos da tela. A nossa nuvem de palavras se assemelha a essa imagem:



Como temos aleatoriedade na posição das palavras na nuvem de palavras, a imagem que aparece em seu colab notebook pode ser diferente da apresentada acima.

### **Nuvem de palavras dos estados no formato do mapa do Brasil**

Nossa nuvem de palavras já ficou bem interessante! Mas você sabia que podemos customizar o nosso objeto WordCloud() adicionando imagem de fundo, máscara, ajuste de cores e muito mais?

Pois bem, é exatamente isso que faremos agora! Para isso vou deixar aqui neste [link](https://cdn3.gnarususercontent.com.br/3057-data-visualization/Atividades/Aula+04/mapa_br.png) o mapa do Brasil do site da [Wikipedia](https://pt.m.wikipedia.org/wiki/Ficheiro:Contorno_do_mapa_do_Brasil.svg) que utilizaremos como espaço para escrever a nossa nuvem de palavras. Salve-a com o nome mapa\_br.png para utilizarmos em nosso exemplo.

Com as bibliotecas e dados já importados, vamos fazer alguns ajustes no código para geração da nuvem de palavras. Primeiro, vamos manter o dataset que vamos utilizar e o dicionário com as frequências (contagens) dos dados. Agora, vamos abrir a imagem com o mapa do Brasil transformando-a em um array numpy. O símbolo de til ( ~ ) serve para inverter os valores do array para que possamos utilizar essa imagem como máscara pros dados serem escritos dentro do mapa.

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from wordcloud import WordCloud

from PIL import Image

# Importando o relatório de vendas

vendas = pd.read\_csv("https://raw.githubusercontent.com/alura-cursos/dataviz-graficos/master/dados/relatorio\_vendas.csv")

# Gerando um dicionário de frequências

dic = dict(vendas["estado"].value\_counts())

# Abrir a imagem com o mapa do Brasil e transformá-lo em um array do numpy

# o sinal til(~) inverte a máscara deixando o espaço dentro do mapa como local de desenho

image\_mask = ~np.array(Image.open('mapa\_br.png'))

Por fim, vamos recriar nosso objeto WordCloud() agora passando alguns parâmetros, como: cor do fundo, largura e altura da imagem, imagem a ser utilizada como máscara, cor do contorno da figura e as cores das palavras. A inserção dos dado no cloud e a personalização da área da figura segue da mesma forma:

# Criar um objeto wordcloud

wordcloud = WordCloud(background\_color = 'white', # cor de fundo

width = image\_mask.shape[1], # largura

height = image\_mask.shape[0], # altura

mask = image\_mask, # imagem

contour\_color = 'white', # cor do contorno

colormap = 'nipy\_spectral\_r') # cor das palavras

# Cria uma wordcloud através do dicionário de frequências

wordcloud.generate\_from\_frequencies(frequencies = dic)

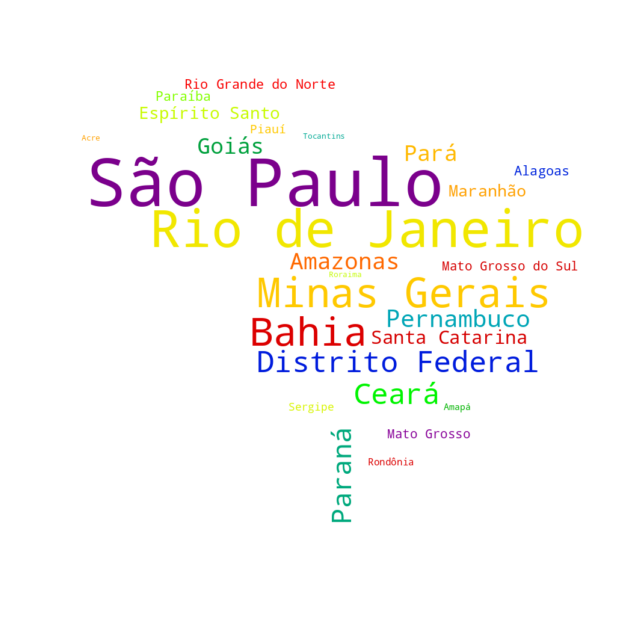
fig, ax = plt.subplots(figsize = (10, 8))

plt.imshow(wordcloud, interpolation = 'bilinear')

ax.axis("off") # retirando bordas e ticks

plt.show()

A nossa nova nuvem de palavras se assemelha a imagem abaixo:



Note que as palavras agora foram ajustadas para que coubessem dentro do formato da imagem do mapa do Brasil. Muito legal, né?

Como temos aleatoriedade na posição das palavras na nuvem de palavras, a imagem que aparece em seu colab notebook pode ser diferente da apresentada acima.

Como sugestão, você pode acessar os seguintes links para se aprofundar na biblioteca da WordCloud:

* [Como fazer uma nuvem de palavras (Bonita) em Python](https://carodias.medium.com/como-fazer-uma-nuvem-de-palavras-bonita-em-python-fec7f9336ab1)
* [Gerar nuvens de palavras de qualquer formato no Python (em ingl](https://www.geeksforgeeks.org/generate-word-clouds-of-any-shape-in-python/)