

## Atividade Prática 2

### Objetivo

Ler, por meio de código python, um dataset (base de dados) de postagens no Facebook de dois presidenciáveis e fazer uma caracterização (análise) inicial dos dados.

### Orientações

Você deverá processar e fazer uma rápida análise (detalhes a seguir) dos arquivos `lula.gz` e `bolsonaro.gz`, que contem as postagens dos respectivos políticos no período de setembro de 2020 a junho de 2021. Note que o dataset está armazenado em arquivos compactados (gz), que são lidos por meio da função `open` da biblioteca `gzip`, que permite a leitura de arquivos compactados. Em cada linha do arquivo há um dicionário (formato “chave”:”valor” conforme visto na última aula) com os detalhes da postagem, incluindo texto (“message”), tipo da postagem (“status\_type”), data e horário de criação (“created\_time”), dentre outros.

Utilize como ponto de partida o código de exemplo presente no arquivo “`read_posts_from_facebook_dataset.py`” para ler a base de dados das postagens no Facebook dos candidatos Lula e Bolsonaro. Note que o código carrega cada linha do arquivo (referente a cada postagem), “transforma” em um dicionário python, acessa dois campos (message e status\_type) e exibe os campos na tela.

Você deverá fazer uma breve caracterização inicial da base de dados. A caracterização nada mais é do que é o processo de analisar e descrever as propriedades e características principais dos dados contidos em uma base. Neste caso, não faremos uma caracterização detalhada, mas obteremos algumas informações que podem nos dar já alguns *insights* sobre os dados. Dentre as análises fornecidas você deverá prover:

- Quantas postagens cada candidato fez?
- Qual foi a palavra mais utilizada pelo candidato nas postagens? Se possível desconsidere palavras como: “que”, “e”, “a”, “da”, etc.
- Qual foi o dia com a maior quantidade de postagens?
- Qual foi postagem feita pelo candidato foi a mais curtida? E a menos curtida?
- Qual foi o tipo de postagem mais utilizado? Vídeo, texto no status --> utilize o campo “status\_type”: “added\_video”, mobile\_status\_update, etc

Cheque quais informações sobre as postagens estão disponíveis no dataset e pense em mais 3 análises (pelo menos) que seriam interessantes de ser feitas. Iremos discutir as análises feitas na próxima aula.