

Para a realização desse trabalho a ideia foi extrair os dados uma semana antes das eleições para fazer a comparação com o resultado das eleições, mas o instituto DataFolha fez somente nos dias 27 e 29 da última semana, então usei esses dois dias na análise de dados.

1. Primeiro passo: pegando os dados do twitter

Para pegar os dados do Twitter utilizando a API python-twitter-v2 , também foi utilizado o snsrape, NLTK, e o pandas .

- O primeiro passo foi a instalação dos módulos necessários e os imports que vão ser utilizados.

```
!pip install python-twitter-v2
!pip install snsrape
!pip install nltk
!pip install pandas
```

```
import snsrape.modules.twitter as dados
import pandas as pd
import datetime
import nltk
```

- Próximo passo foi criar dois Arrays que vão ser utilizados para acumular os tweets buscados.

```
lista_tweets_bolsonaro = []
lista_tweets_lula = []
```

- Próximo passo foi buscar os tweets que mencionam o Bolsonaro dos dias 23 a 29 de outubro, foi extraído 500 tweets de cada dia.

```
#coleta dados do dia 23 de outubro
for i,tweet in enumerate(dados.TwitterSearchScraper('Bolsonaro since:2022-10-22 until:2022-10-24').get_items()):
    if i>500:
        break
    lista_tweets_bolsonaro.append([tweet.date, tweet.id, tweet.content, tweet.user.username])
```

```
#coleta dados do dia 24 de outubro
for i,tweet in enumerate(dados.TwitterSearchScraper('Bolsonaro since:2022-10-23 until:2022-10-25').get_items()):
    if i>500:
        break
    lista_tweets_bolsonaro.append([tweet.date, tweet.id, tweet.content, tweet.user.username])
```

```
#coleta dados do dia 25 de outubro
for i,tweet in enumerate(dados.TwitterSearchScraper('Bolsonaro since:2022-10-24 until:2022-10-26').get_items()):
    if i>500:
        break
    lista_tweets_bolsonaro.append([tweet.date, tweet.id, tweet.content, tweet.user.username])
```

```
#coleta dados do dia 26 de outubro
for i,tweet in enumerate(dados.TwitterSearchScraper('Bolsonaro since:2022-10-25 until:2022-10-27').get_items()):
    if i>500:
        break
    lista_tweets_bolsonaro.append([tweet.date, tweet.id, tweet.content, tweet.user.username])
```

```
#coleta dados do dia 27 de outubro
for i,tweet in enumerate(dados.TwitterSearchScraper('Bolsonaro since:2022-10-26 until:2022-10-28').get_items()):
    if i>500:
        break
    lista_tweets_bolsonaro.append([tweet.date, tweet.id, tweet.content, tweet.user.username])
```

```
#coleta dados do dia 28 de outubro
for i,tweet in enumerate(dados.TwitterSearchScraper('Bolsonaro since:2022-10-27 until:2022-10-29').get_items()):
    if i>500:
        break
    lista_tweets_bolsonaro.append([tweet.date, tweet.id, tweet.content, tweet.user.username])
```

```
#coleta dados do dia 29 de outubro
for i,tweet in enumerate(dados.TwitterSearchScraper('Bolsonaro since:2022-10-28 until:2022-10-30').get_items()):
    if i>500:
        break
    lista_tweets_bolsonaro.append([tweet.date, tweet.id, tweet.content, tweet.user.username])
```

O mesmo código acima foi utilizado para o Lula, mudando a menção para Lula e utilizando o array `lista_twittes_lula`.

- Após isso foi criado um DataFrame com os dados da extração dos twittes.

```
: #Montando o DataFrame com os dados.  
df_bolsonaro = pd.DataFrame(lista_twittes_bolsonaro, columns=['Datetime', 'Tweet Id', 'Text', 'Username'])  
df_lula = pd.DataFrame(lista_twittes_lula, columns=['Datetime', 'Tweet Id', 'Text', 'Username'])
```

- Próximo passo será instalar alguns módulos novos e importá-los.

```
!pip install textblob  
!pip install unidecode
```

```
from textblob import TextBlob  
from unidecode import unidecode
```

O textblob conta o número de palavras que exprimem um sentimento positivo e negativo em cada frase, e o unidecode faz a formatação de dados de texto em Unicode para ASCII.

pyhtobn

- Após isso vamos instalar o módulo que é uma biblioteca python gratuita e ilimitada que implementou a API do Google Translate e importar-lá.

```
!pip install googletrans
```

```
from googletrans import Translator  
translator = Translator()
```

- Próximo passo é criar um contador de sentimentos positivos, negativos e totais.

```
#Contadores de sentimentos positivos e negativos  
numPos = 0  
numNeg = 0  
total = 0
```

- Próximo passo é aplicar unicode nos tweets do Bolsonaro.

```
#Aplicando unidecode aos textos do Tweet  
textPT = df_bolsonaro['Text'].apply(unidecode)  
display(textPT)
```

- Agora vamos obter os sentimentos de cada tweet do Bolsonaro.

```
sentimento_bolsonaro = []  
#Traduzindo para o Inglês  
for x in textPT:  
    textEN = translator.translate(x) #traduzindo  
    sentimento = TextBlob(textEN.text) #obtendo o sentimento  
    sentimento_bolsonaro.append(sentimento.polarity) #adicionando a uma lista  
display(sentimento_bolsonaro)
```

- Após isso vamos armazenar os dados dos sentimentos em um arquivo excel.

```
#Criando DataFrame no formato CSV  
df_bolsonaro['Datetime'] = df_bolsonaro['Datetime'].apply(lambda a: pd.to_datetime(a).date())  
df_sentimentol = pd.DataFrame.from_dict(data={'Data':df_bolsonaro['Datetime'],'Bolsonaro Sentimento':sentimento_bolsonaro},  
orient='columns')  
df_sentimentol.to_excel('sentimento_bolsonaro.xlsx', index=False)
```

O mesmo processo acima foi utilizado para fazer a contagem dos twittes do lula e criar um arquivo excel com os dados dos sentimentos.

2. Segundo passo: extrair os dados da pesquisa eleitoral.

Vamos extrair os dados das pesquisas eleitorais do site base do dados, eles disponibilizam um data leak para consulta, segue o link: https://basedosdados.org/dataset/br-poder360-pesquisas?bdm_table=microdados, eles também mantêm os dados atualizados e tratados.

Crie um outro arquivo separado para realizar extração da pesquisa eleitoral.

- Para iniciar vamos fazer a instalação do módulo deles.

```
!pip install basedosdados
```

- Vamos fazer os imports necessários.

```
import basedosdados as bd
import pandas as pd
```

- Agora para realizar o próximo passo será necessário criar uma conta no Google cloud e criar um projeto, após isso pegue seu project ID e substitua no trecho de código abaixo onde está o "projeto_id", vamos atribuir a tabela ao DataFrame.

```
# Para carregar o dado direto no pandas
df = bd.read_table(dataset_id='br_poder360_pesquisas',
table_id='microdados',
billing_project_id="projeto_id")
```

- Podemos ver as informações das pesquisas eleitorais.

```
In [4]: df.head()
```

```
Out[4]:
```

	id_pesquisa	ano	sigla_uf	nome_municipio	cargo	data	data_referencia	instituto	contratante	orgao_registro	...	tipo	turno	tipo_voto	id_ce
0	8023	2006	None	None	presidente	2006-09-19	18-19.set.06	Datafolha	None	None	...	estimulada	1	votos totais	2
1	8023	2006	None	None	presidente	2006-09-19	18-19.set.06	Datafolha	None	None	...	estimulada	1	votos totais	2
2	8023	2006	None	None	presidente	2006-09-19	18-19.set.06	Datafolha	None	None	...	estimulada	1	votos totais	2
3	8023	2006	None	None	presidente	2006-09-19	18-19.set.06	Datafolha	None	None	...	estimulada	1	votos totais	2
4	8023	2006	None	None	presidente	2006-09-19	18-19.set.06	Datafolha	None	None	...	estimulada	1	votos totais	2

5 rows x 24 columns

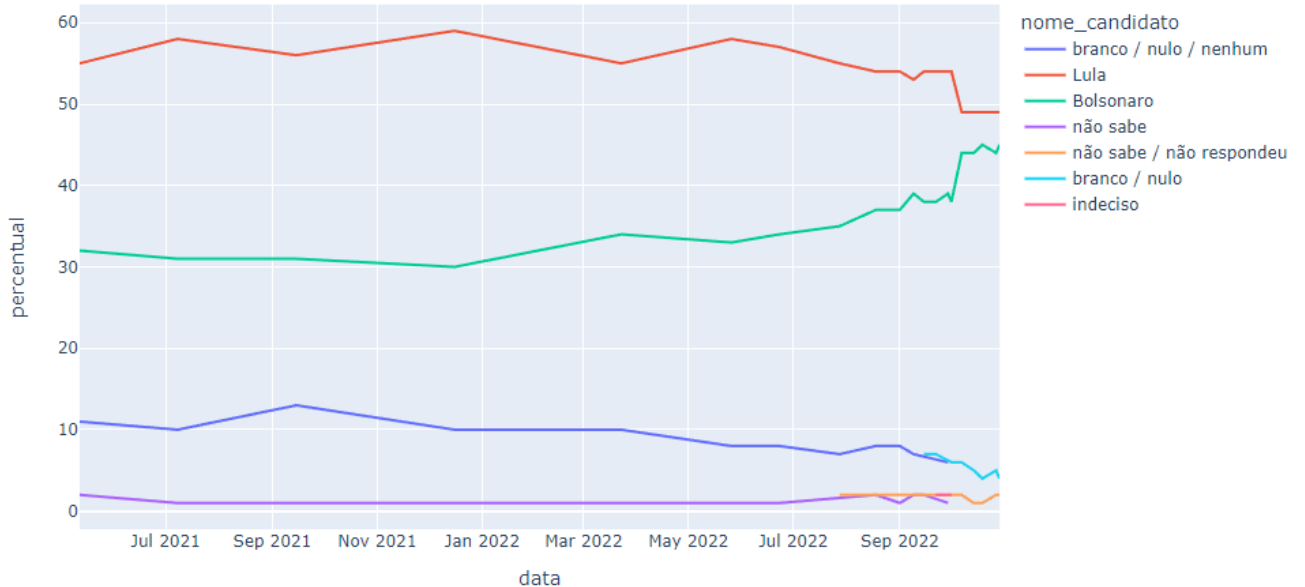
- Vamos fazer um filtro para a proposta do trabalho.

```
In [5]: pesquisas_2022 = df[(df.instituto == 'Datafolha') #essa linha filtra pelo instituto DataFolha.
& (df.ano == 2022) #essa linha filtra pelo ano.
& (df.cargo == 'presidente') #essa linha filtra pelo cargo.
& (df.sigla_uf.isnull()) #essa linha filtra pra pegar os dados de todo o Brasil, sem distinção de UF.
& (df.tipo == 'estimulada') #essa linha filtra pra pegar pesquisa estimulada.
& (df.turno == 2) #essa linha filtra pelo turno.
& (df.descricao_cenario.str.contains('cenário 1')==True) #essa linha filtra para o cenário 1
].sort_values(by='data') # essa linha ordena a tabela pela data.
```

- Após o filtro podemos utilizar o Plotly para fazer o gráfico do DataFrame filtrado.

```
In [7]: pd.options.plotting.backend = "plotly"
```

```
In [8]: pesquisas_2022.plot(x='data', y='percentual', color='nome_candidato')
```



- Por fim vamos armazenar os dados filtrados em um arquivo excel

```
#Criando DataFrame no formato CSV
pesquisas_2022['data'] = pesquisas_2022['data'].apply(lambda a: pd.to_datetime(a).date())
pesquisas_2022.to_excel('pesquisas_2022.xlsx', index=False)
```

3. Utilização dos arquivos excel gerados.

Foi gerado três arquivos excel na pasta do projeto, um com os dados do sentimento dos tweets em relação ao Bolsonaro, outro com os dados do sentimento dos tweets em relação ao Lula, e por último o da pesquisa eleitoral do instituto DataFolha.



Esses arquivos foram carregados no power BI para serem tratados e por fim ser gerado um painel com os dados de sentimentos e com os dados da pesquisa eleitoral.