FeelingAnalysisTweet

Work for information retrieval discipline

-> Implement (in python) a collector that requests and collects data through an API such as: Twitter API, Github, Linkedin, Facebook. And it extracts some information, presenting a report of conclusions about the extracted data.

-> The collection / extraction must have an objective and the same must be clear. The objective should not be just to collect, but to present some analysis with some justification for this and the relevance of that analysis.

-> Perform unit tests for each function implemented

-> Generating a Jupyter report there will be:
- Contextualization of the definition of the collection objective

- Justification and relevance of this collection (why is it important?)

- Presentation of the analysis and its conclusions. You must use APIs to display graphics such as: Pandas, Seaborn, matplotlib.

The implementation code must be readable:

-> Name of variables / functions that make their function clear in code

-> Well commented code

-> Code well divided into functions / classes

Introduction

"In the midst of the Covid-19 pandemic, hundreds of millions of people worldwide and the economies of more than 190 countries are trying to adapt to the current reality and curb the threats of recession. The guidelines of the World Health Organization (WHO) on the subject are objective: to contain the spread of the disease it is necessary to promote social isolation and reinforce hygiene habits. " - UFJF

The chaotic political and social scenario in Brazil and in the world, calls for a general analysis of society that is increasingly adept at social networks. In this way, analyzing this environment brings us a vision of what society is thinking about a certain subject. This work aims to research on the current situation of the President of Brazil, Jair Bolsonaro, in this way a sentiment analyzer was created that receives data from twiiter and makes its analysis verifying the amount, in that time interval, which is the position of the twiiter in relation to president.

Abstract

At this moment, defending science and knowledge as strategic tools to face this pandemic and also as a condition for the country's development is part of the defense of life and of our public teaching and research institutions. Using society analyzes makes us prepared and better understand how the population is reacting, so this work aims to analyze society using the social network twitter in the last 30 days to know the position of people who use the hashtag bolsonaro.

Technologies

In this first stage of the project, the name, messages and the device being sent by twiiter were collected. To analyze the messages, a dataframe system was used that was manipulated by api pandas, the analysis of feelings was made by the api TextBlob that makes the analysis for texts in English, in this way to translate the texts in Portuguese, the api googletrans was used and in order to have no unwanted characters in the text, the unidecode.

```
In [1]:
from src.collectorTweet import tweet

newtt = tweet("#bolsonaro",100,"pt")
newtt.feelingAnalizer(60)
```

Looking for 100 twitters about the word #bolsonaro, 60 days ago, wait ...

Analyzing this nation's feelings ...
Analyzing the data, please wait ...
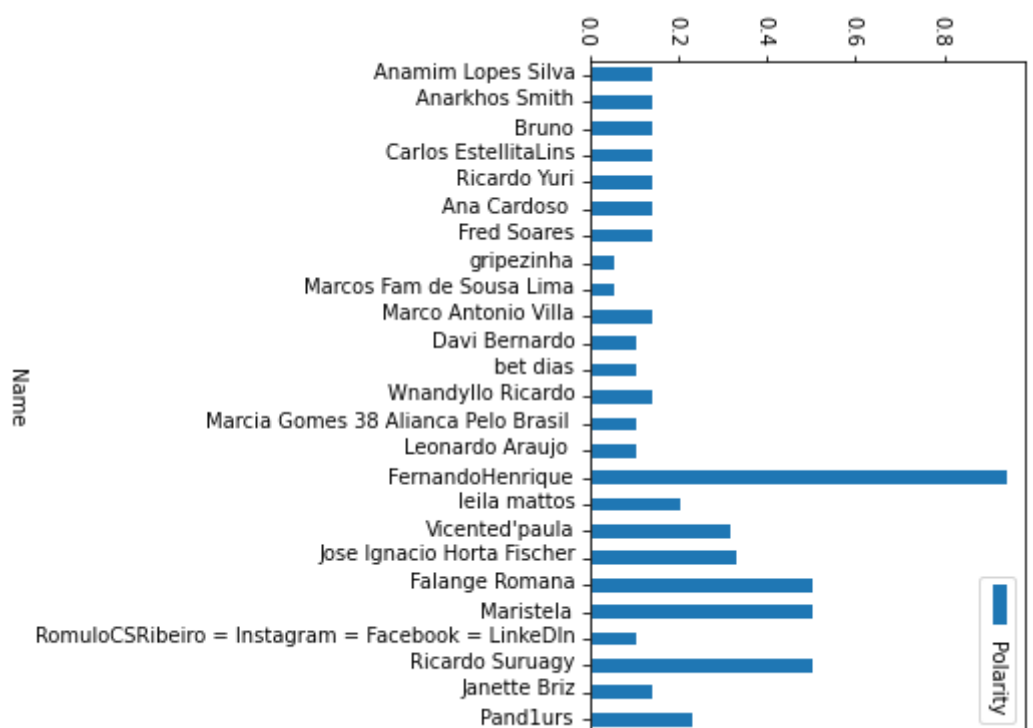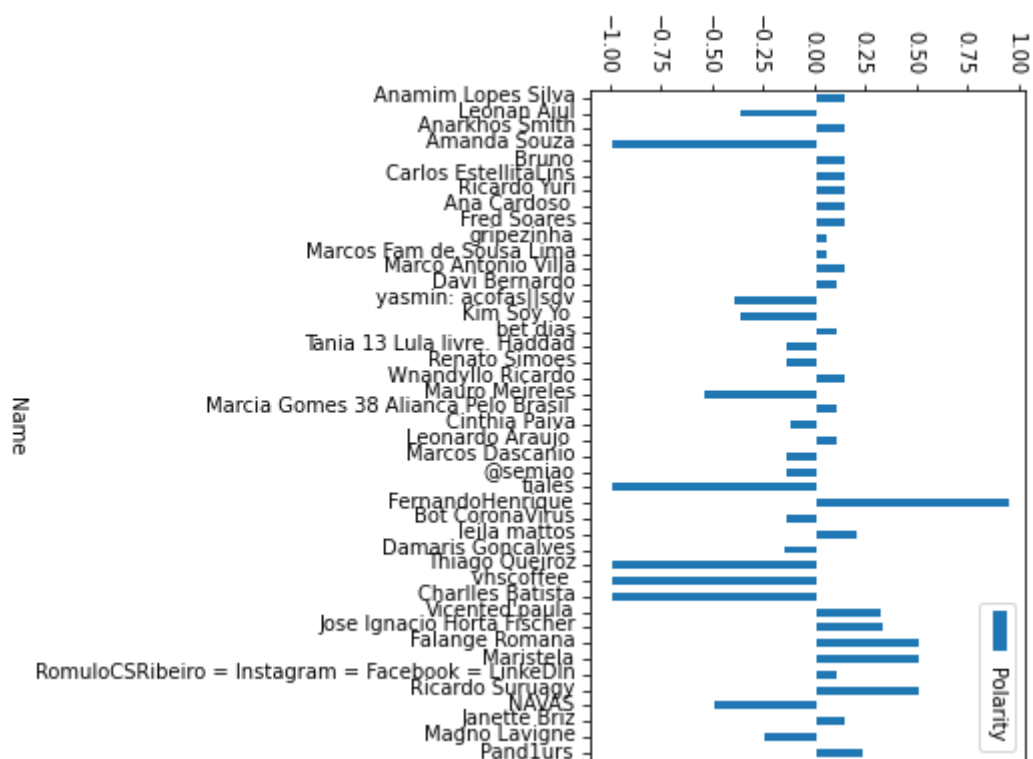Printing general data

43 analyzed data.
With the average -0.07002407798919427, we see that twitter are hating the situation of this research.
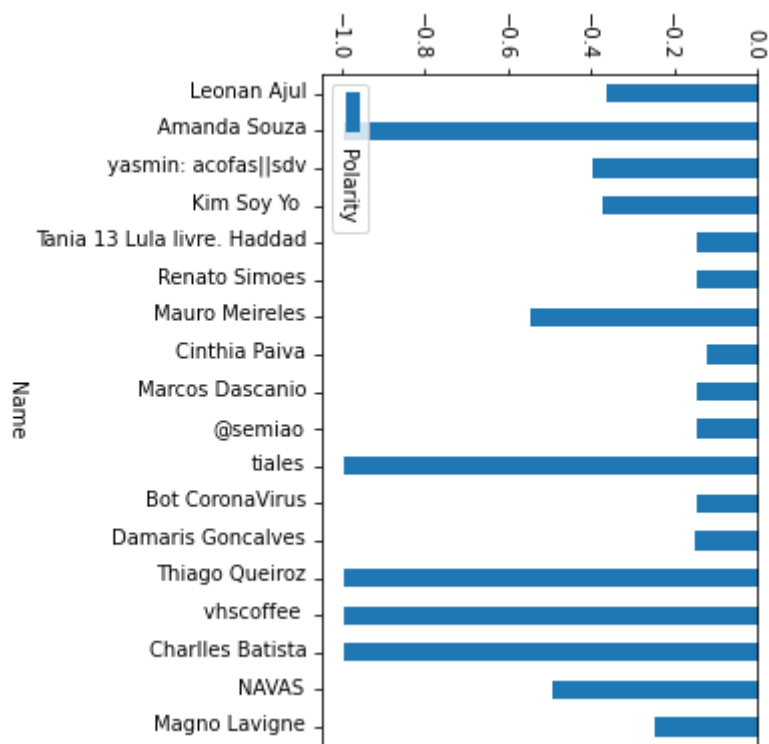Positive numbers :
{'Amount': 25, 'Max': 0.9375, 'Min': 0.05000000000000002, 'Average': 0.2185808080808081}
Negative Number:
{'Amount': 18, 'Max': -0.125, 'Min': -1.0, 'Average': -0.47086419753086417}

Results

The results were not as assertive as expected, on average, 60% of the twitters managed to be analyzed, and of these, a smaller portion still had high accuracy, this is mainly due to the translation that reduced the reliability of the analysis. In the future, we intend to use other methods of analysis such as wastson, or even a method specifically trained specifically for the Brazilian language, in addition, using other aspects such as emojis, other networks and covering other words would be efficient to improve the result.