



SENTIMENT ANALYSIS ASSIGNMENT

Dabzat Samira, Diallo Djenabou Diari, Haroun Boukar Khadidja



DATA PRE-PROCESSING

1) *Translation :*

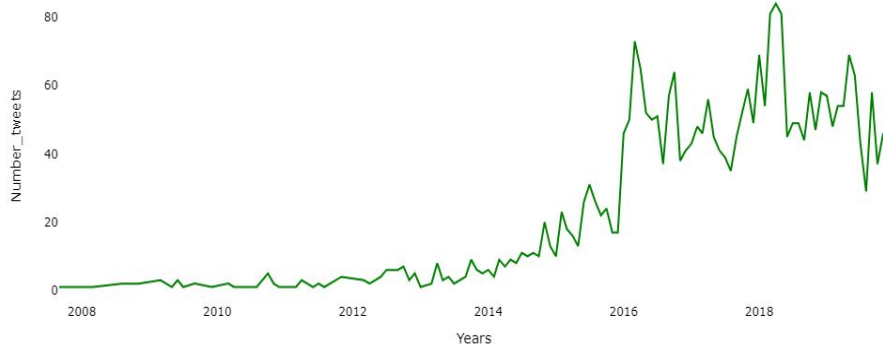
- Library googletrans' Translator()

2) *Cleaning:*

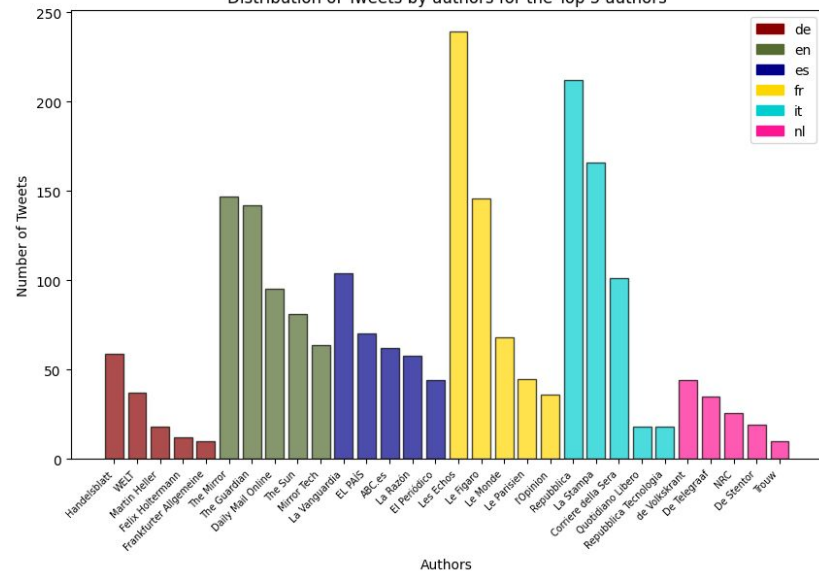
- Extraction of the # in a new column
- Lower, removal of all the unwanted (emojis, usernames...)

DATA EXPLORATION

L'évolution des Tweets au fil du temps



Distribution of Tweets by authors for the Top 5 authors







DATA ANALYSIS

1) *Name entity recognition :*

index	Entity Type	Probability
0	PERSON	0.212953
1	ORG	0.199232
2	CARDINAL	0.156970
3	GPE	0.138310
4	DATE	0.097695
5	NORP	0.082327
6	ORDINAL	0.053238
7	TIME	0.013172
8	LOC	0.012075
9	QUANTITY	0.008782
10	PRODUCT	0.008782
11	FAC	0.008233
12	MONEY	0.004940
13	EVENT	0.002195
14	PERCENT	0.001098
15	WORK_OF_ART	



DATA ANALYSIS

2) *Sentiment analysis:*

- Pipeline From Roberta Base sentiment
 - sentiment for each tweets by the top 5 authors
 - average sentiment by author
 - mapped the sentiment (1 : positive, 0 : neutral, -1 : negative)

SENTIMENT ANALYSIS

