

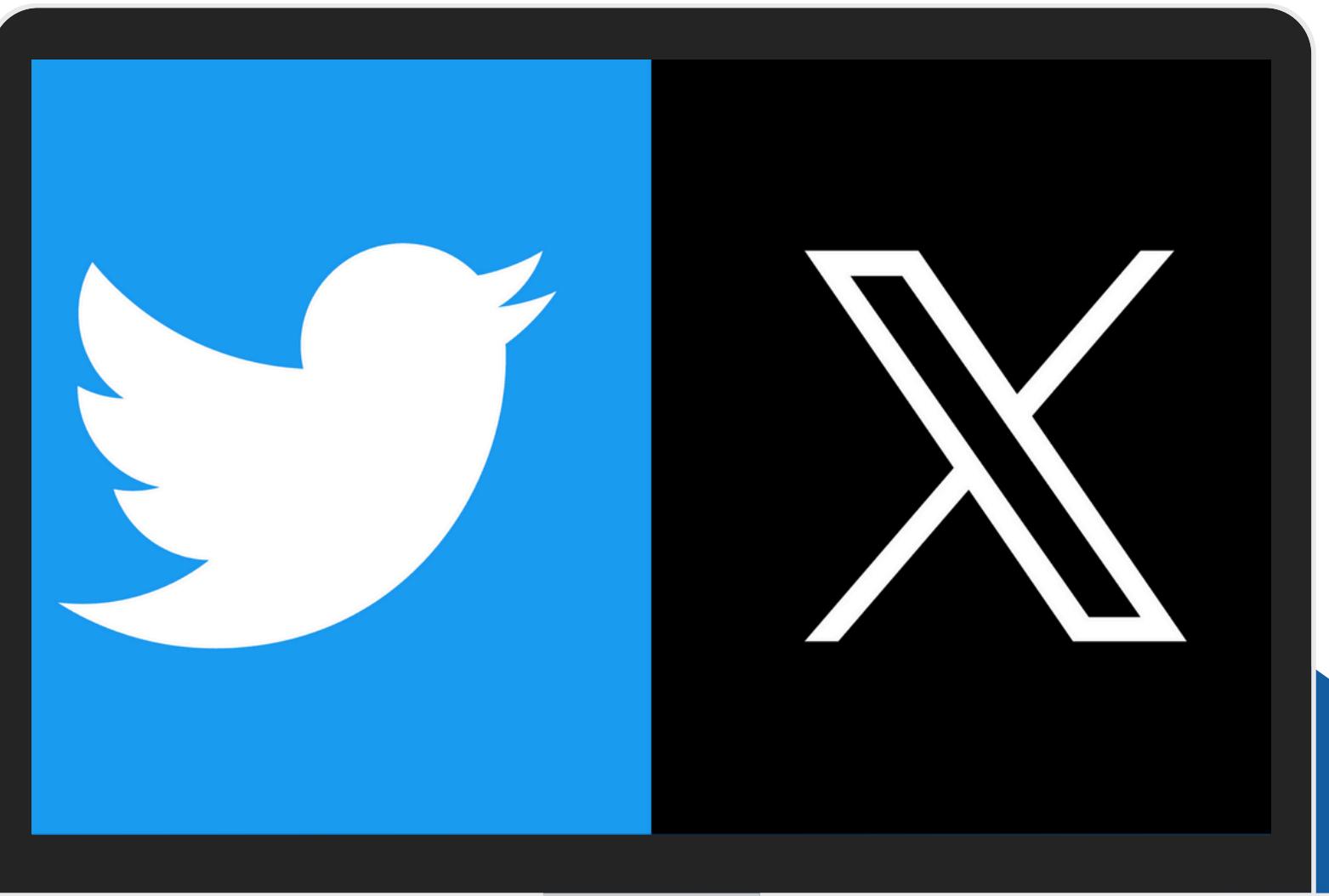


Tweets controversies

Directed by Jonathan DUCKES, Audric GIRONDIN, Maéva
MAÏO, Houria SAYAH, Alya ZOUZOU



Janssen
Horizon



INTRODUCTION

- **Controversie** : an exchange of divergent opinions highlighting contradictory points of view
- **Nonpharmacologic treatment** : preventive measures and care strategies that are not linked to drugs



Emmanuel Macron

@EmmanuelMacron

To protect and slow the spread of Covid-19: from Monday until further notice, nurseries, schools, colleges and universities will be closed.

8:58 PM · March 12, 2020

19,9 k Retweets 2 748 Quotes 67,4 k Likes 880 Bookmarks

How data visualisations can contribute to a better understanding of the factors behind health controversies on Twitter ?

Table of contents

▶ Project Management	3
▶ Data processing	5
▶ Visual Exploration and Processing of Tweets	7
▶ Conclusions	14
▶ Future prospects	16



Project management

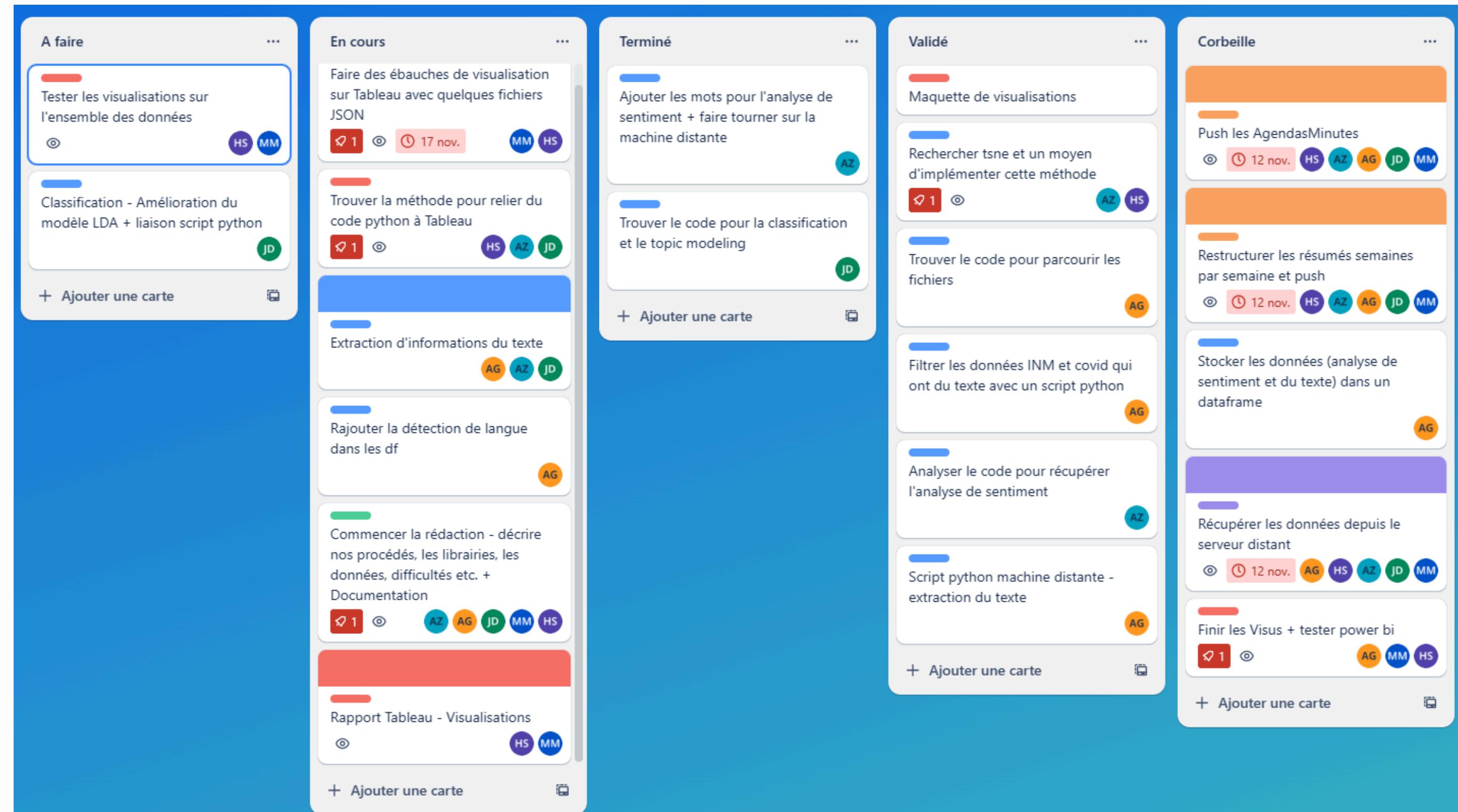
Project Manager

Task allocation

Task Monitoring

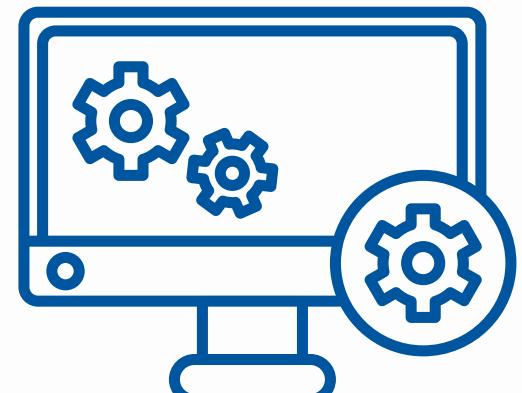
Meeting with the Supervisor

Meeting Team



Used tech

```
11 var b = count_array_gen();
12 parseInt(liczenie().unique);
13 var c = array_bez_powt(); a = " ",d = parseInt($("#limit_val"),
14 function("LIMIT_total:" + d);
15 function("rand:" + f);
16 d < f && (f = d, function("check rand\u000f3\u000f3rand: " + f +
17 var h = [], d = d - f, e;
18 if (0 < c.length) {
19   for (var g = 0;g < c.length;g++) {
20     e = indexOf_Array(b, c[g]), -1 < e && b.splice(e, 1);
21   }
22   for (g = 0;g < c.length;g++) {
23     b.unshift({use_class:"parameter", word:c[g]});
24   }
25   void function(b) {
26     var e = parent();
27     e = parent();
28     e = parent();
29     e = parent();
30     e = parent();
31     e = parent();
32     for (c = 0;c < d && c < b.length;c++) {
```



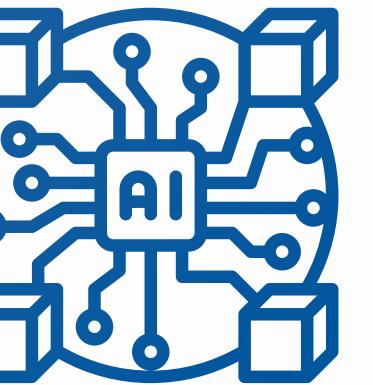
Softwares

Github

Tableau

Tableau Prep

MiniConda



Algorithms

Community Louvain

ForceAtlas2

Tokenizer

Transformers



Libraries

Gensim

JSON

Matplotlib

MultiProcessing

NLTK

Pyvis

TextBlob

Hugging Face

LangDetect

MemoryProfiler

NetworkX

Pandas

Spicy

Data processing



Raw data

Number of tweets : 5 486 603

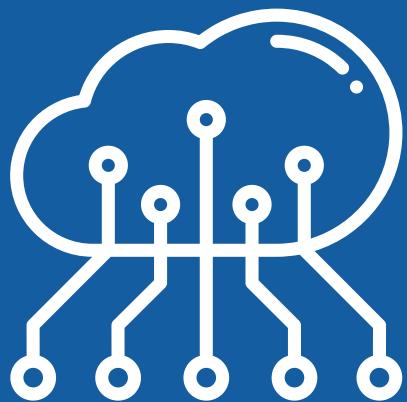
Size : 23 G



Data processing

- **Data Retrieval**
- **Sorting and Organizing**
- Data**
- **Python Filtering Script**
- **Processing and Analyzing**
- Tweets**
- **Preparation for**
- Visualization**





Our data

ID	Nb retweet	Nb like	Nb réponses	Nb citations	Hashtags	Texte	Mots
1147982997838340099	5	17	2	0.0	['#FASTING', '#PANCREATIC', '#CANCER']	#Fasting reduces intestinal toxicity from radi...	['Fasting', 'reduces', 'intestinal', 'toxicity...']
1072525224640425985	0	0	0	0.0	['#ALKALINE', '#FOOD', '#HEALER', '#NATURALREM...']	https://t.co/TanAjknG1L\nTips for the best tim...	['https', 't.co/TanAjknG1L', 'Tips', 'for', 't...']
1405348861078478852	0	0	0	0.0	['#COFFEE', '#CANCER', '#FASTING', '#INTERMITT...']	☕It's been testified that #coffee intake is b...	['☕It', '"', 's', 'been', 'testified', 'that']
1407724898550755329	0	0	0	0.0	['#REMINDERPOST', '#HINDU', '#FESTIVALS', '#PU...']	Tomorrow is Purnima!\n24 June 2021\n#reminderp...	['Tomorrow', 'is', 'Purnima', '24', 'June', '2...']
627028927458988032	0	0	0	0.0	['#WANNAKNOW', '#FASTING', '#HEALTH', '#CANCER']	#WannaKnow Does #fasting help shrink a tumor o...	['WannaKnow', 'Does', 'fasting', 'help', 'shri...']

Shape : 249 456, 17

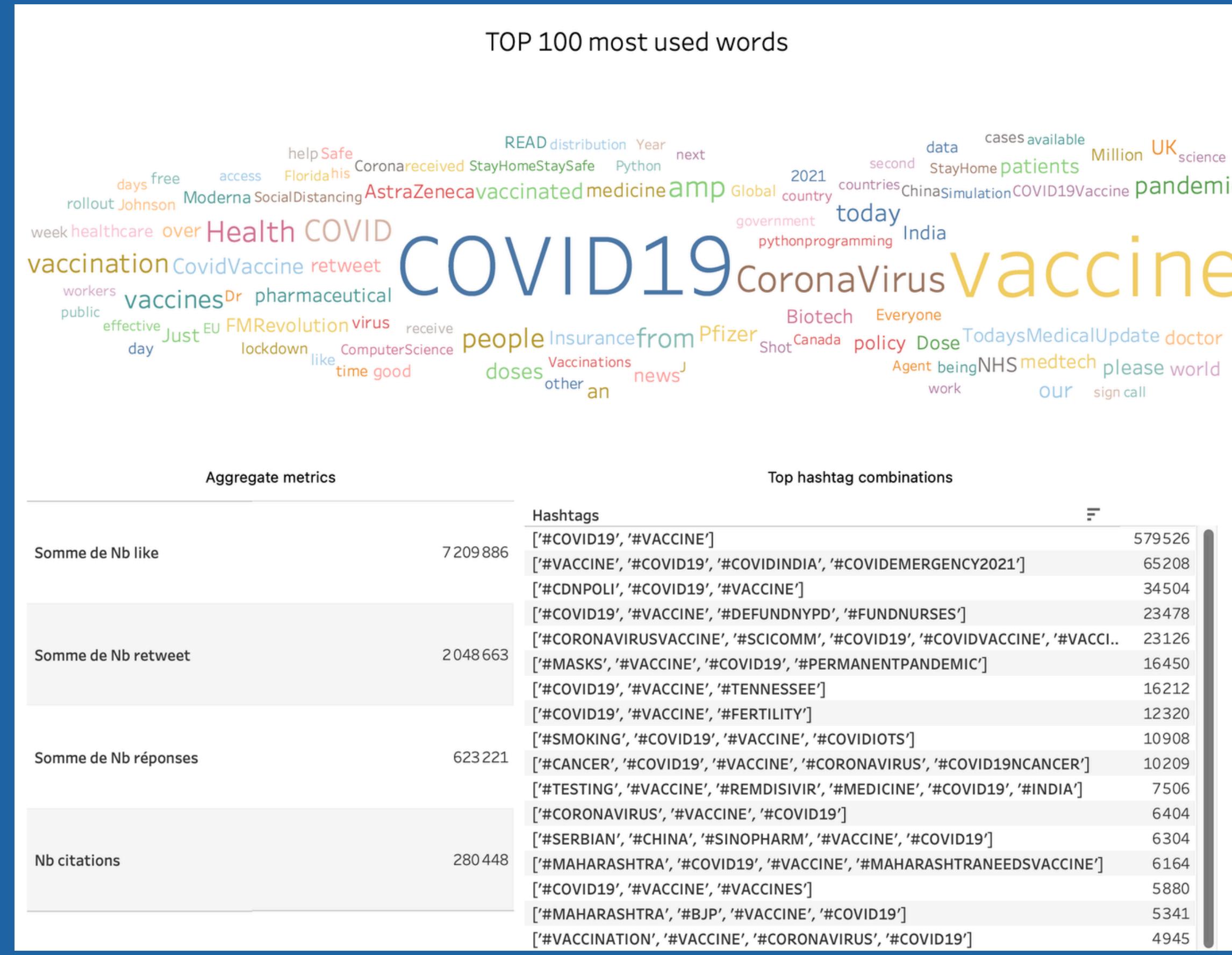
Likes : 7 209 886

Retweets : 2 048 663

Responses : 623 221

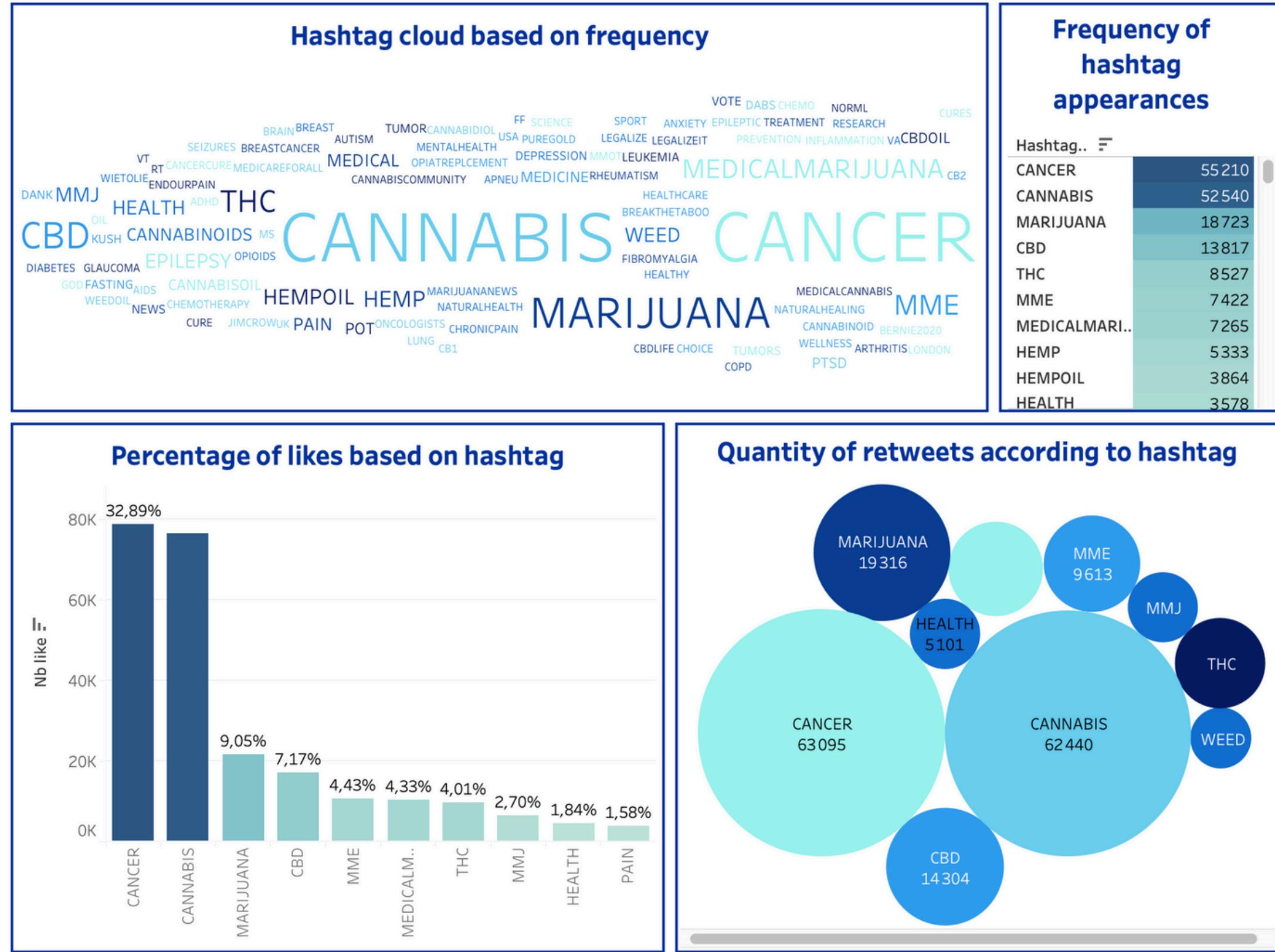
Citations : 280 448

Visual Exploration and Processing of Tweets



Visualization

NMI Analysis



Sentiment & Emotion Analysis with Hugging Face

RoBERTa

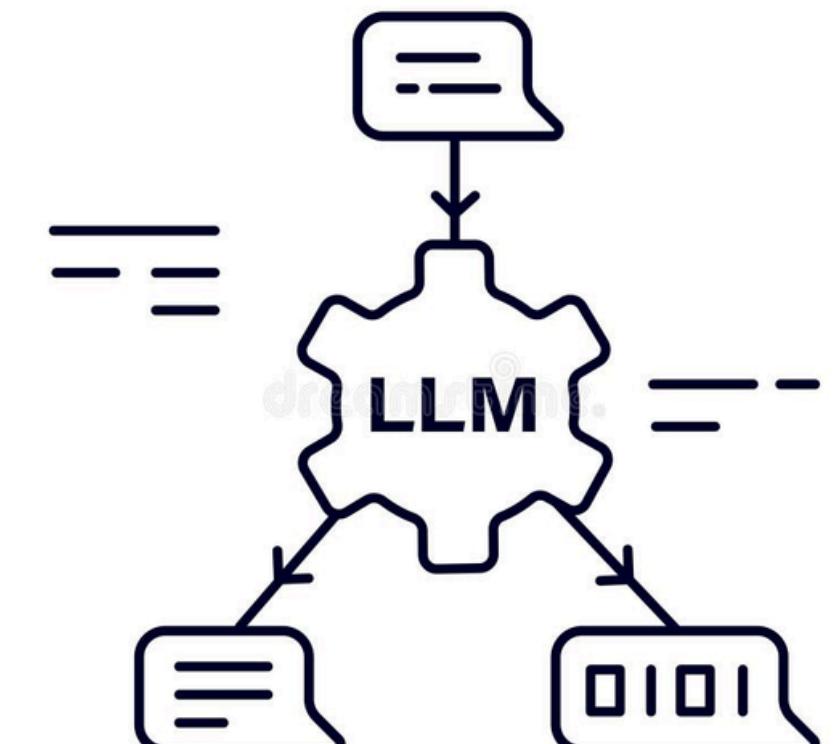
- Bidirectional Understanding
- Expanded Training Dataset
- Longer Training Period
- Dynamic Masking



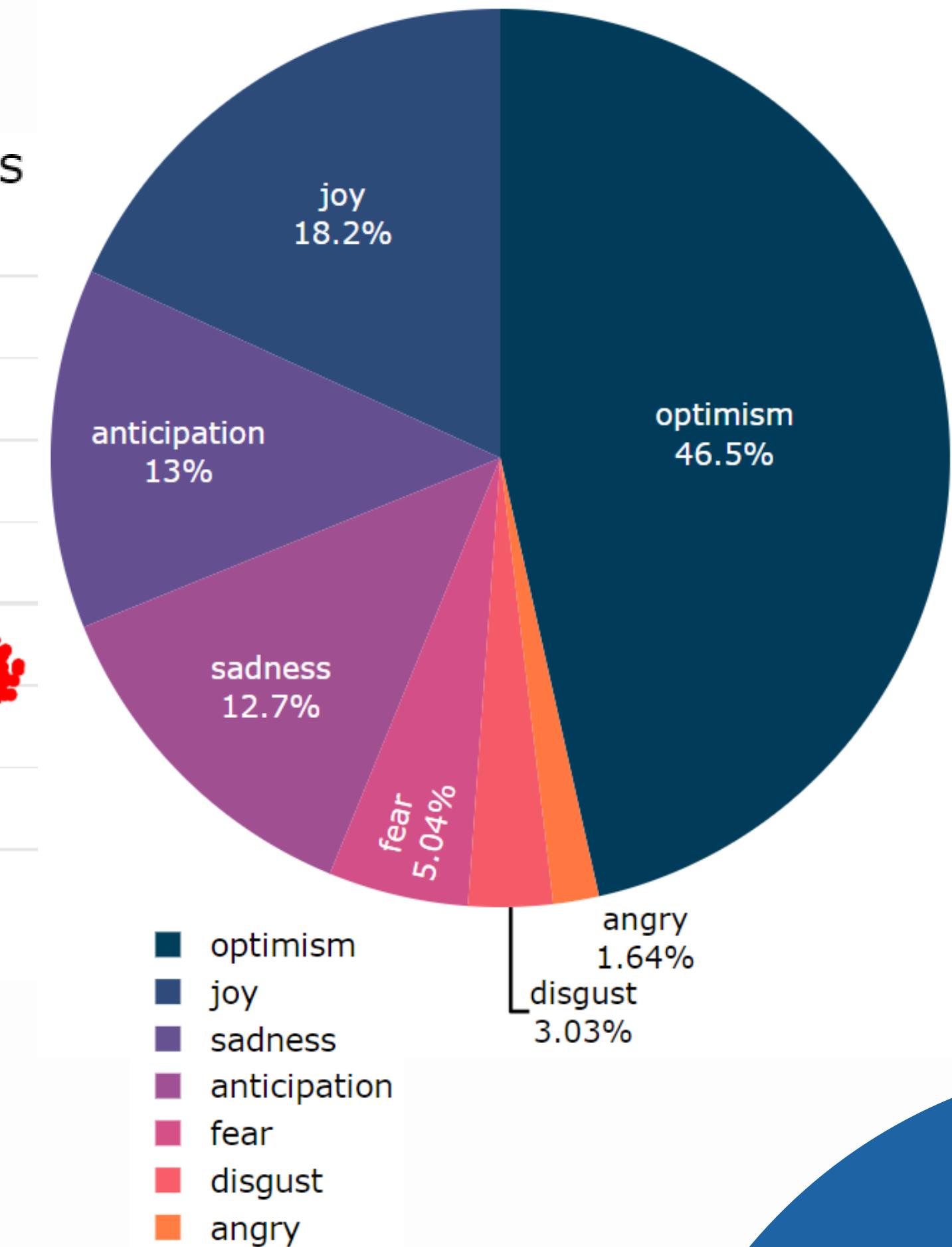
Cardiff NLP

How ?

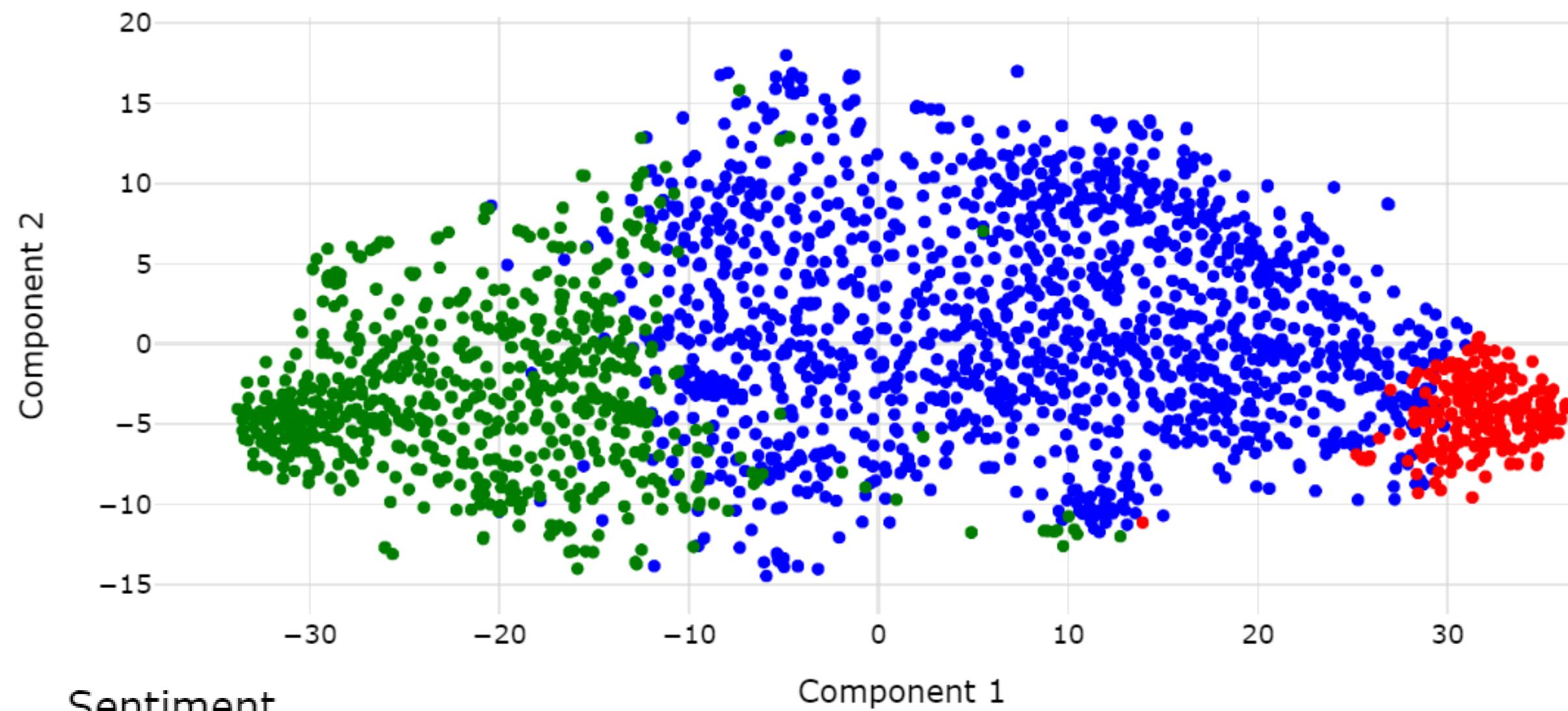
 Meta



Proportion of Each Emotion



t-SNE Visualization of Feature Spaces for Sentiment Analysis



Sentiment

- neutral
- pos
- neg

Topic Modeling - LDA

- Topic Modeling : determines cluster words for a set of documents
- David Blei developped the Latent Dirichlet Allocation model
- The LDA model identifies clusters that can be interpreted to infer key topics.

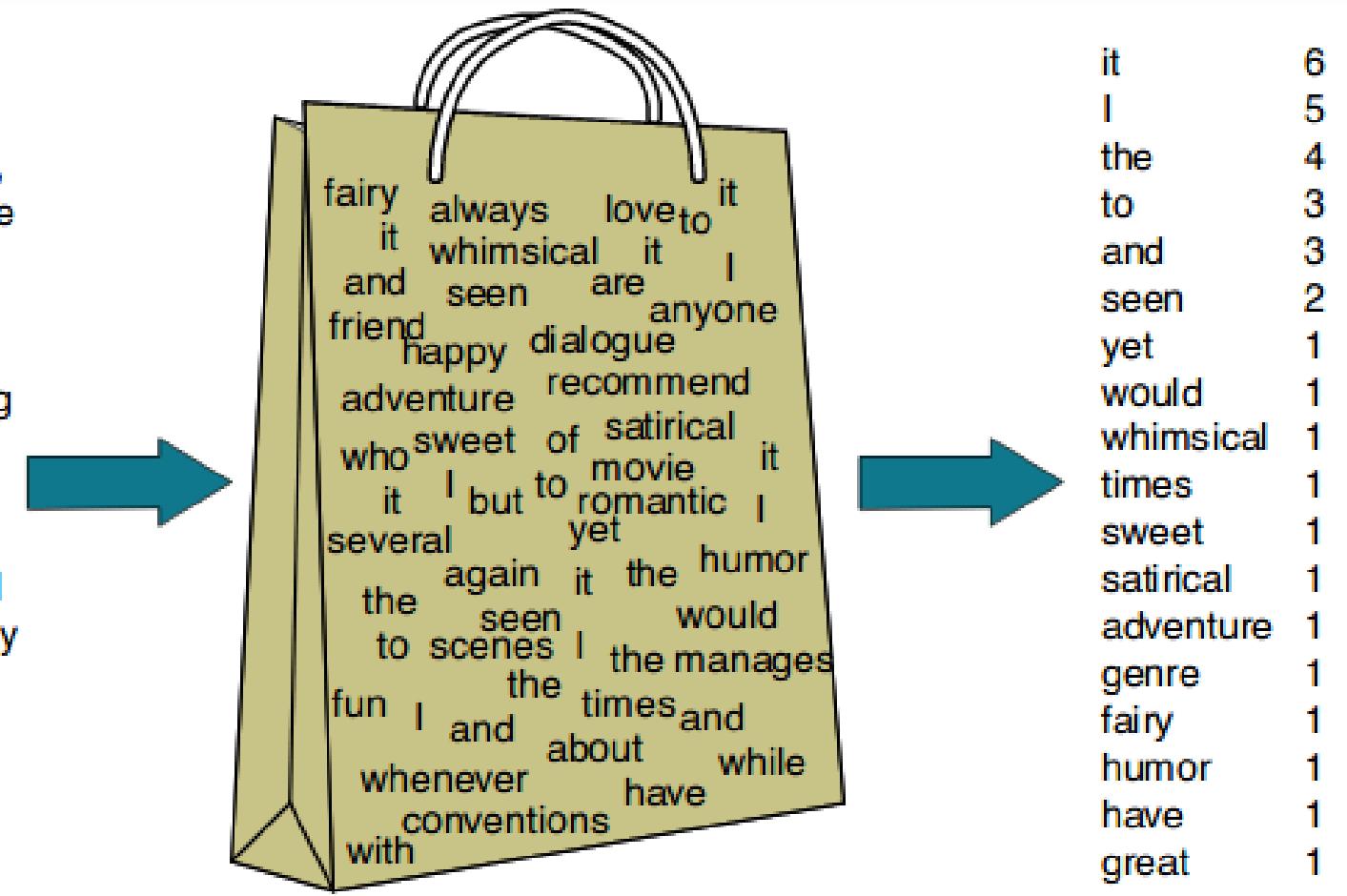
studying
studies
study

stemming

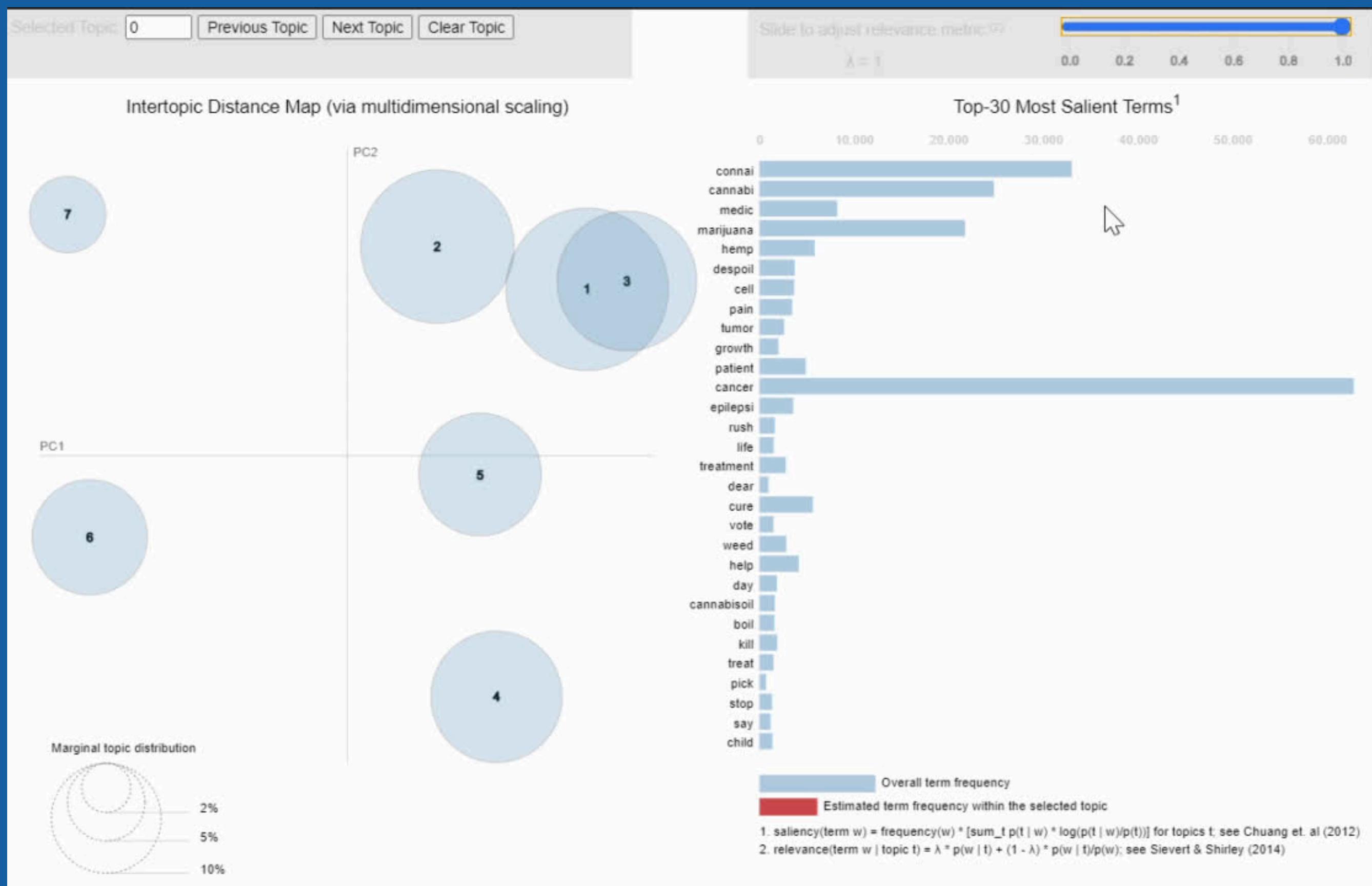
→

studi
studi
studi

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!

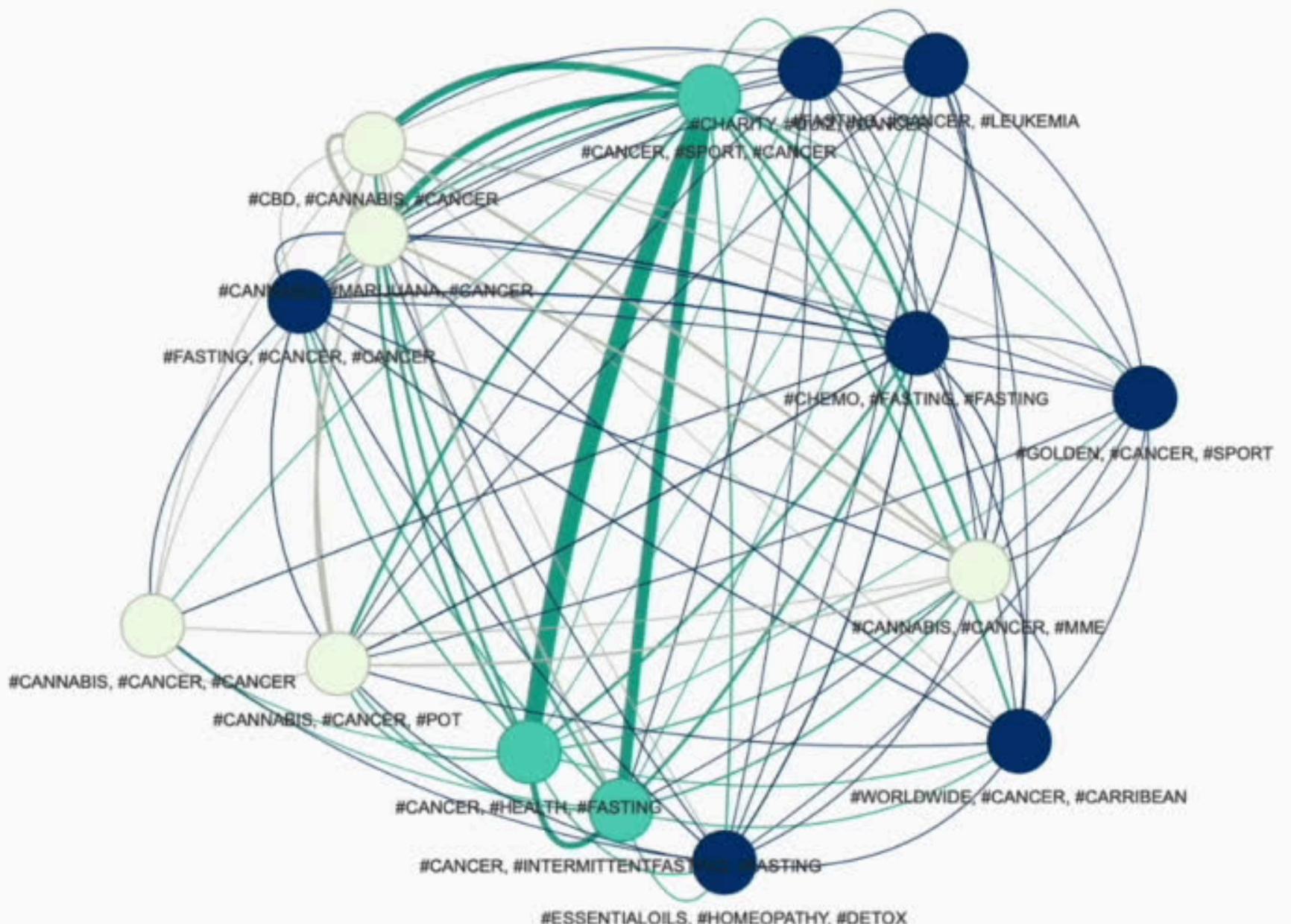


LDA VISUALIZATION



Community Detection in the Hashtag Cluster Graph

- Clusters
- Co-occurrences
- Community



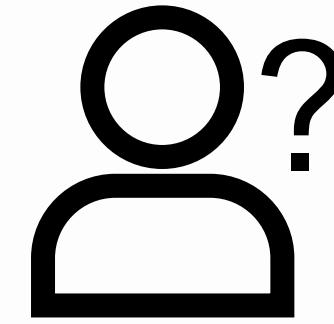
Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, 9(6), e98679.

Que et al. Scalable Community Detection with the Louvain Algorithm, In Proceedings of the 2015 IEEE International Parallel and Distributed Processing Symposium, Hyderabad, India, 2015, 28-37

CONCLUSIONS

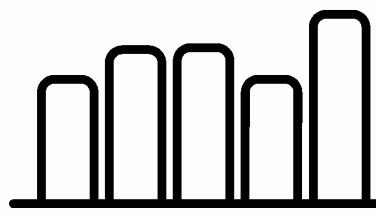
- Diversity of health topics discussed
 - Emotions
 - Discussion topics
 - Community clusters
- } Significant contributions to online debates

Difficulties



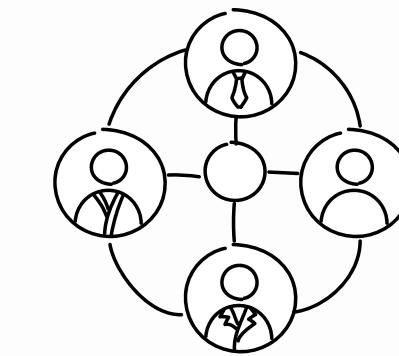
Expectations

- To define our objectives
- Problematic
- Actions to be carried out



Mass of Data

- Everybody
- Long waiting time : execution of scripts, visualizations loading ...



Coordination

- Coordinate the work of everyone

Future prospects

01

Link between the different parts

02

Controversy index

03

Increase coherence score and add a label to each topic

04

Hate speech detection

05

Establish everything based on other datasets





THANK YOU!

Without forgetting our supervisors :
Sandra BRINGAY and Maximilien SERVAJEAN

