

# La France n'a pas dit son dernier mot

Word Frequency Analysis (Twitter and Amazon)

# Agenda

- Team
- Context
- Topic
- Data Sources
- Workflow
- Challenges faced
- Achievements
- Insights
- Conclusions
- Going Further



# The Dream Team

Aurélie

Experience in text Analysis



Wonders why do people purchase this book.

Paulina

# Context

Eric Zemmour latest book “La France n’a pas dit son dernier mot” is one of the best seller in France: 8th at a national level, 1st in the Political Section in Amazon.fr and among the first 5 in other bookstores.

Our question is WHY?



# Research Question

What is the word frequency of comments on Amazon regarding the Eric Zemmour latest book compared with the word frequency of hashtag's #lafrancenapasditsonderniermot comments ?

# Data sources

API Twitter  
Whole Twit

## Web Scraping AMAZON



MarcTrivor



**Un excellent ouvrage. A lire absolument.**

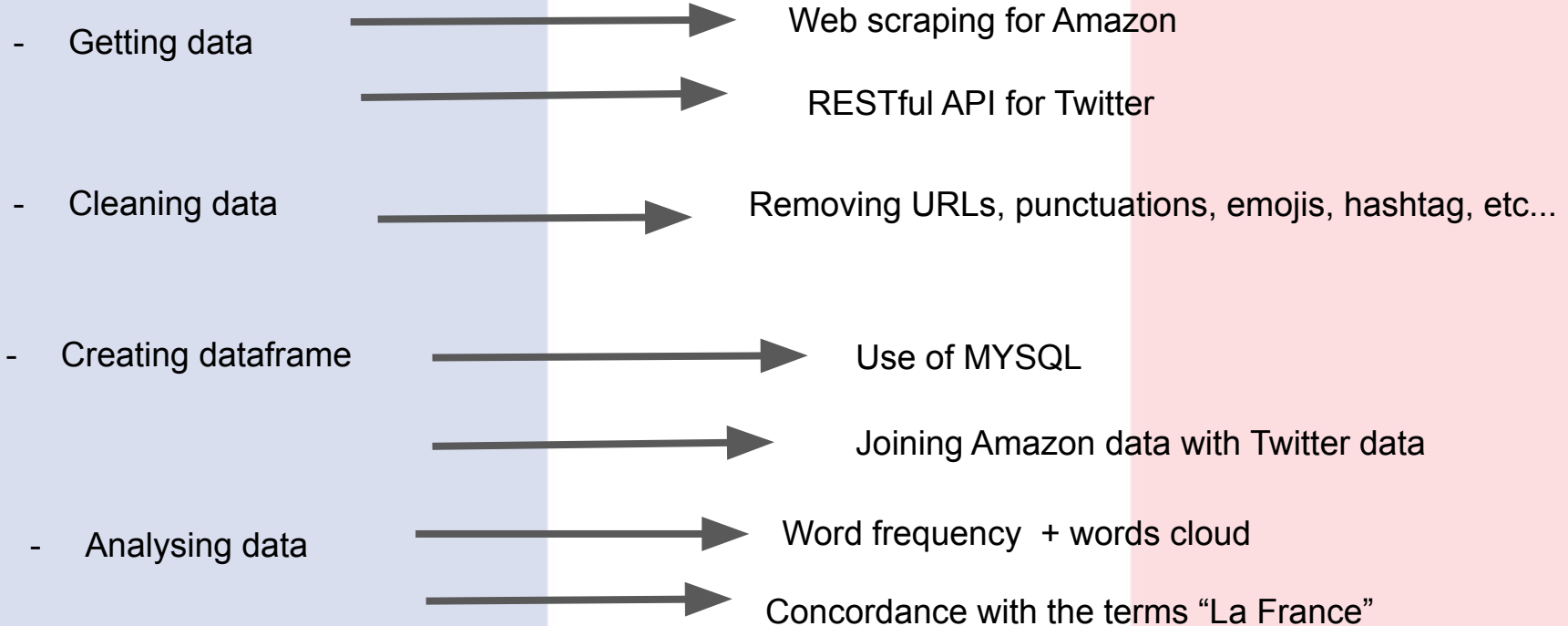
Commenté en France le 16 septembre 2021

Un excellent ouvrage. A lire absolument. Eric Zemmour raconte avec une précision et une fluidité sans précédent l'histoire de la France moderne. J'ai lu tous les autres livres de ce même auteur et je ne peux qu'approuver ses dires. La France est en danger de mort, il faut impérativement la sauver. L'union fait la force. Vive la France et Vive Monsieur Z.

712 personnes ont trouvé cela utile



# Workflow: the Four steps



# Challenges faced

## Tweets

Creation and authorization of a Twitter project

- Limited data to analyse.
- Many metacharacters to clean.

## Common Challenges

- Not enough time to explore interesting methods to make text analysis.
- Not enough material to make a relevant analysis.
- Difficulties to clean the text with a words list, specially the characters in French.

## Commentaires Amazon

- How to webscrape without an API?
- FNAC banned us.
- AMAZON:poursuite judiciaire?

# Achievements

- We consolidate a dataframe with all the text scrapped from Twitter and Amazon.
- We became experts in Twitter Web Scraping with an API.
- We learned several methods in order to display data analysis (word cloud, word frequency, concordance, ...)  
However we did not have the time these methods properly.



# Insights

```
[92]: # Une autre méthode "concordance" : montre les occurences d'un mot dans son contexte
print("Exemples d'occurences du terme 'france' :")
text2.concordance("france")
print('\n')
```

Exemples d'occurences du terme 'python' :  
Displaying 11 of 11 matches:  
ité sans précédent l'histoire de la france moderne. j'ai lu tous les autres li  
ne peux qu'approuver ses dires. la france est en danger de mort, il faut impé  
ver. l'union fait la force. vive la france et vive monsieur z. livre de l'anné  
rt. constat lucide sur l'état de la france et notamment la dérive tragique de  
s. une lueur d'espoir cependant, la france n'a pas dit son dernier mot. génie  
uis des décennies et qui ont mis la france dans l'état dans laquelle on la ret  
de celui qui peut encore sauver la france du chaos qui s'est installé inexora  
uxième génération. effectivement la france n'a pas dit son dernier mot. mais q  
présentera à la présidentielle. la france en a besoin, et oui, la france n'a  
. la france en a besoin, et oui, la france n'a pas dit son dernier mot. le tru  
t sa vie privée. ;) urgence pour la france le z vient de démontrer une nouve

```
# Une autre méthode "concordance" : montre les occurences d'un mot dans son contexte
print("Exemples d'occurences du terme 'vérité' :")
text2.concordance("raison")
print('\n')
```

Exemples d'occurences du terme 'vérité' :  
Displaying 2 of 2 matches:  
zemmour a raison zemour a raison sur tout il dit la  
zemmour a raison zemour a raison sur tout il dit la verite ..... un

```
# Create and generate a word cloud image:
wordcloud = WordCloud().generate(w_joined)

# Display the generated image:
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



Words	Frequency
zemmour	26
a	20
raison	2
zemour	2
sur	17
tout	17
il	9
dit	9
la	75
verite	1
un	33
livre	25
exceptionnel	2
quel	1
courage,	1
quelle	3
plume	4
!	17

# Conclusions

- Time-consuming analysis
- Word Cleaning is mandatory in order to get correct outputs during the analysis.
- Lemmatization and Sentiment Analysis could have been interesting to complete this work because we have an heterogenous corpus, but with words that can be grouped.
- Even if the analysis made don't allow to came to a clear conclusion, we found out that most of the words used, either in Twitter or in Amazon, are positive terms.
- Also, the word “France” is often used in contexts that praise nationalism, unit and remembrance of past glory.



# Going Further

- Code simplification for the Amazon Part.
- Comparison with other data sources (Reddit, Facebook, comments in articles, FNAC).
- Analysis of Eric Zemmour's discourses
- Lemmatization analysis of words in comments
- Sentiment analysis of words in comments



Thank you

