Analysing newspaper articles using Python

Cláudia Silvestre[1], Júlia Leitão de Barros[1], Carina Ferreira[1] and Pedro Frazão
[1] Escola Superior de Comunicação Social - IPL

Abbreviated abstract: This work aims to analyse how the pandemic was covered by the Portuguese media. Giving special emphasis to what the reader feels about the news and which areas/themes were the most affected by the news. We started with a content analysis of the news stories. Then, to analyze the database and identify the most frequent words, we built a dictionary in Python and finally, we verified that there were significant differences between sentiments by means of hypothesis testing.

Related publications:
– Hyejin, P., Biddix, P., & Park, H. (2021). "Discussion, news information, and research sharing on social media at the onset of Covid-19". Profesional de la información, 30(4), 67-77.
– Silvestre, C. e Frazão, P. (2021). COVID-19 news a way to promote statistics classes. Proceedings of ICERI2021 - 14th annual International Conference of Education, Research and Innovation.

4th Conference on
Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

ESCS
ESCOLA SUPERIOR
DE COMUNICAÇÃO SOCIAL

Problem, Data

- **Research Hypotheses**
  - What themes are covered by the news about the pandemic?
    Economy / Politics / Education / Society / Culture / Technology / Sports / Health
  - What feelings do they cause the reader?
    Safety / Fear / Warning / Doubt / Hope
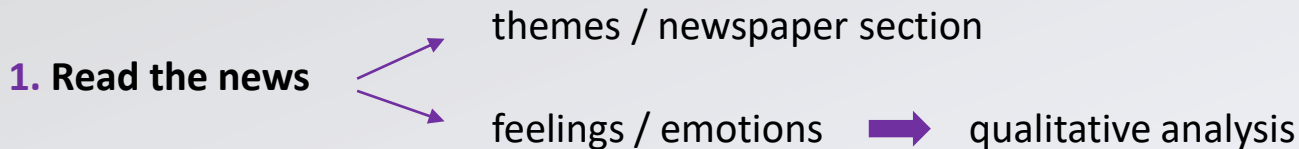 - Are there any words in the headlines that are more associated with certain feelings?

- **Collect data**
  - daily newspapers: Diário de Notícias and Público + weekly: Expresso
  - from January to March, 2021

- **Dataset**
  - Clean data
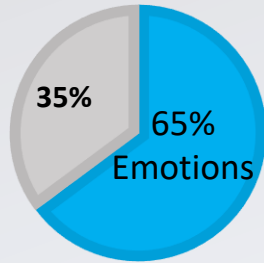 - Prepare data for analysis : analyze the news ➡ complement the database

Methods

1. Read the news

themes / newspaper section

feelings / emotions ➡ qualitative analysis

2. Analyze the database

- To identify the most frequent words in the titles we used **Python** ➡ created a **dictionary**

- **Chi-squared** test of independence

Results and Conclusions

2020 News

23%

77% Nacional

35%

65% Emotions

Python dictionary

- Pandemic (**13 %**)

- Safety and Fear: Covid
- Warning, Doubt and Hope: Vaccines

Chi-squared test

➢ **Themes**

National   International

Health   Vaccine   Technology

Warning Fear   Hope