

# Hespress News Report



# Table of content

- Introduction
- Insights about data by asking and answering questions.
- Conclusion

---

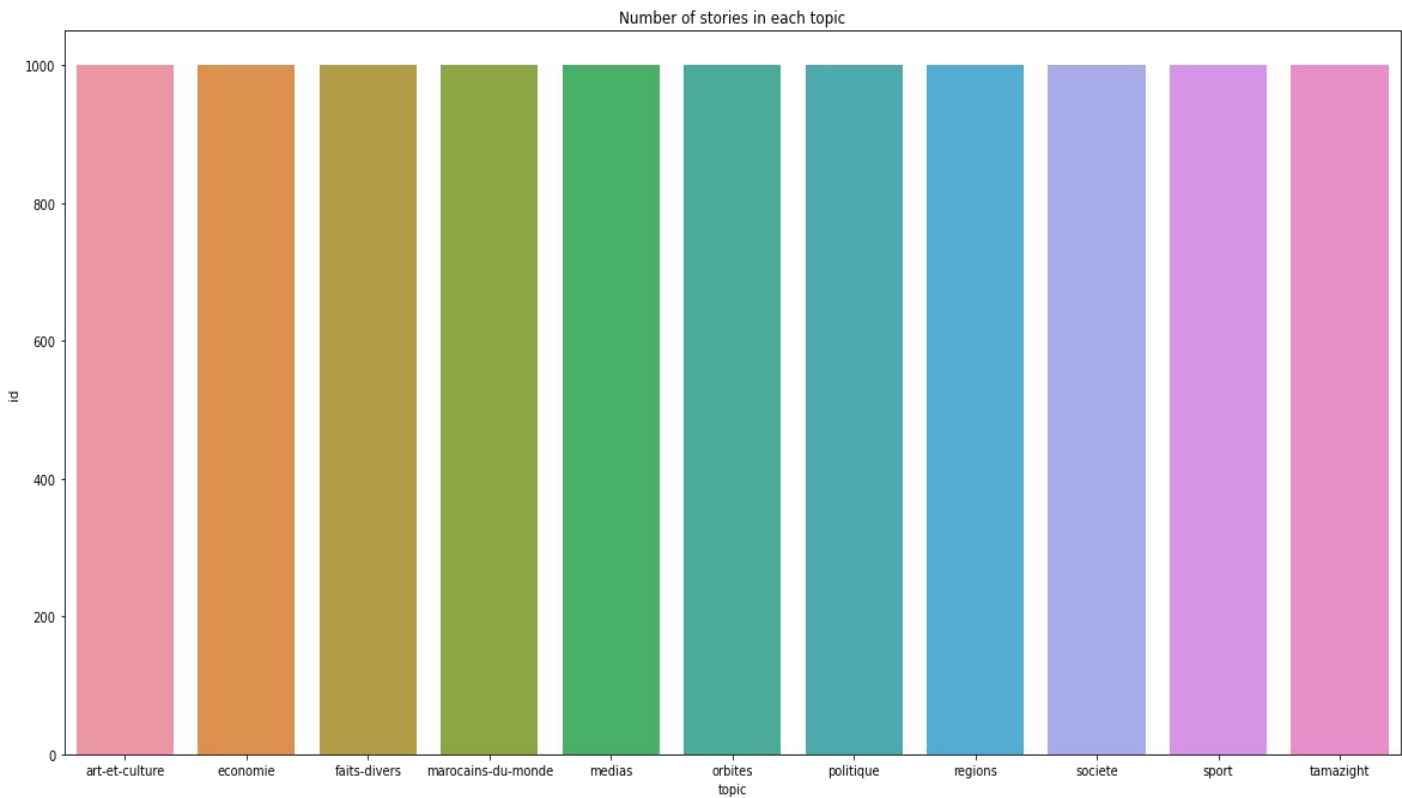
# Introduction

- Hespress is a Moroccan news website that provides news and information about Morocco and the world to Arabic-speaking audiences.
- It was launched in 2007 and has become one of the most popular news websites in Morocco, with millions of monthly visitors.
- Hespress covers a wide range of topics, including politics, society, economy, sports, culture, and technology, among others.



Insights about data by asking and  
answering questions.

# Q1 : What is the number of examples per class?



- **As shown , dataset is balanced and each class has the same number of stories.**

**Q2 : What are top frequent n-grams generally and per class?**

	5-gram	Count
0	،في تصريح لجريدة هسبريس الإلكترونية	512
1	تصريح لجريدة هسبريس الإلكترونية، أن	329
2	الخارجية والتعاون الإفريقي والمغاربة المقيمين	243
3	الشؤون الخارجية والتعاون الإفريقي والمغاربة	240
4	الوطنية والتكوين المهني والتعليم العالي	195

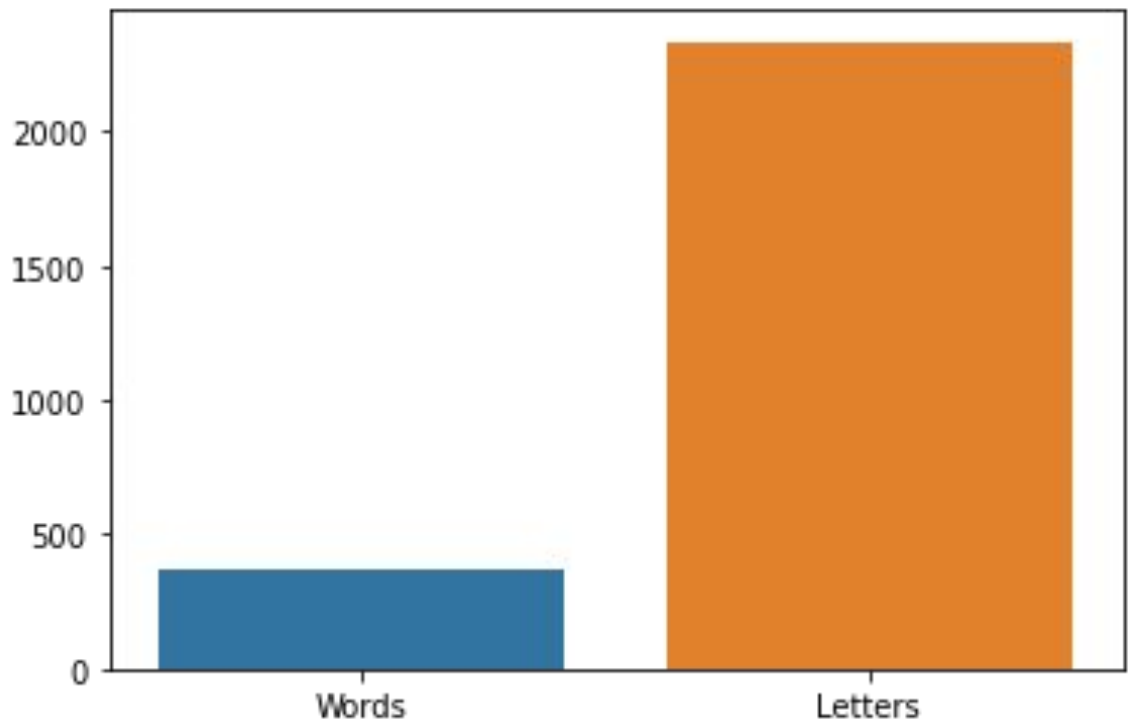
- This table shows the most frequent five 5-grams in stories in general.

	5-gram	Count	class
0	القانون التنظيمي لتنفيذ الطابع الرسمي	92	tamazight
1	الجامعة الملكية المغربية لكرة القدم	58	sport
2	،في تصريح لجريدة هسبريس الإلكترونية	126	societe
3	،في تصريح لجريدة هسبريس الإلكترونية	45	regions
4	الشؤون الخارجية والتعاون الإفريقي والمغاربة	126	politique
5	النبي صلى الله عليه وسلم	49	orbites
6	قراءة مواد بعض الجرائد الورقية	54	medias
7	المنتدبة المكلفة بالمغاربة المقيمين بالخارج	68	marocains-du-monde
8	،تحت إشراف النيابة العامة المختصة	134	faits-divers
9	،وزير الاقتصاد والمالية وإصلاح الإدارة	53	economie
10	،في تصريح لجريدة هسبريس الإلكترونية	36	art-et-culture

- This table shows the most frequent 5-grams per class.



**Q3 : What is the mean length of examples in words and letters?**



- **As shown:**

**1. The mean length of words per story is 371 word .**

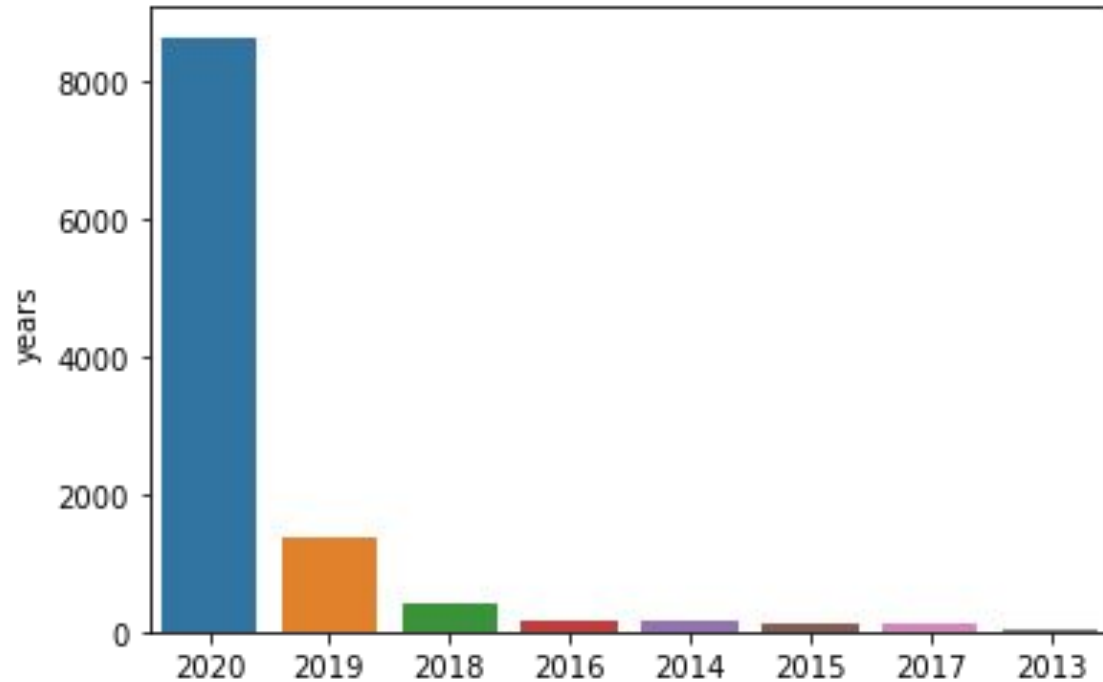
**2. The mean length of letters per story is 2334 letter.**

Q4 : What is the most frequent author name in each class?

	author	count
topic		
art-et-culture	*يوسف صدوق من الرياض	433
economie	هسبريس- محمد بلقاسم	489
faits-divers	ياسين أوشين	259
marocains-du-monde	يونس إيجيري من هلسنكي	372
medias	يوسف لهاتلي من باريس	550
orbites	ياسين عددان	437
politique	ياسين أوشين من خنيفرة	441
regions	ياسين عددان (صور: أحمد بنسماويل)	189
societe	ياسين عددان	523
sport	أوعنا بلعيد	400
tamazight	ياسين كوسكار	166

- As shown ,the most frequent author per class.

**Q5 : What is the most common year in the dataset?**



- As shown, the most common year in the dataset is 2020.
-

# Conclusion

- Dataset is balanced and each class has the same number of stories.
- The most frequent 5-grams in stories in general is repeated 512 times.
- The most frequent 5-grams per class is repeated 134 at faits-divers class.
- The mean length of words per story is 371 word .
- The mean length of letters per story is 2334 letter.
- The most frequent author per class is Youssef Lahlali from Paris with 550 times at medias class.
- The most common year in the dataset is 2020.