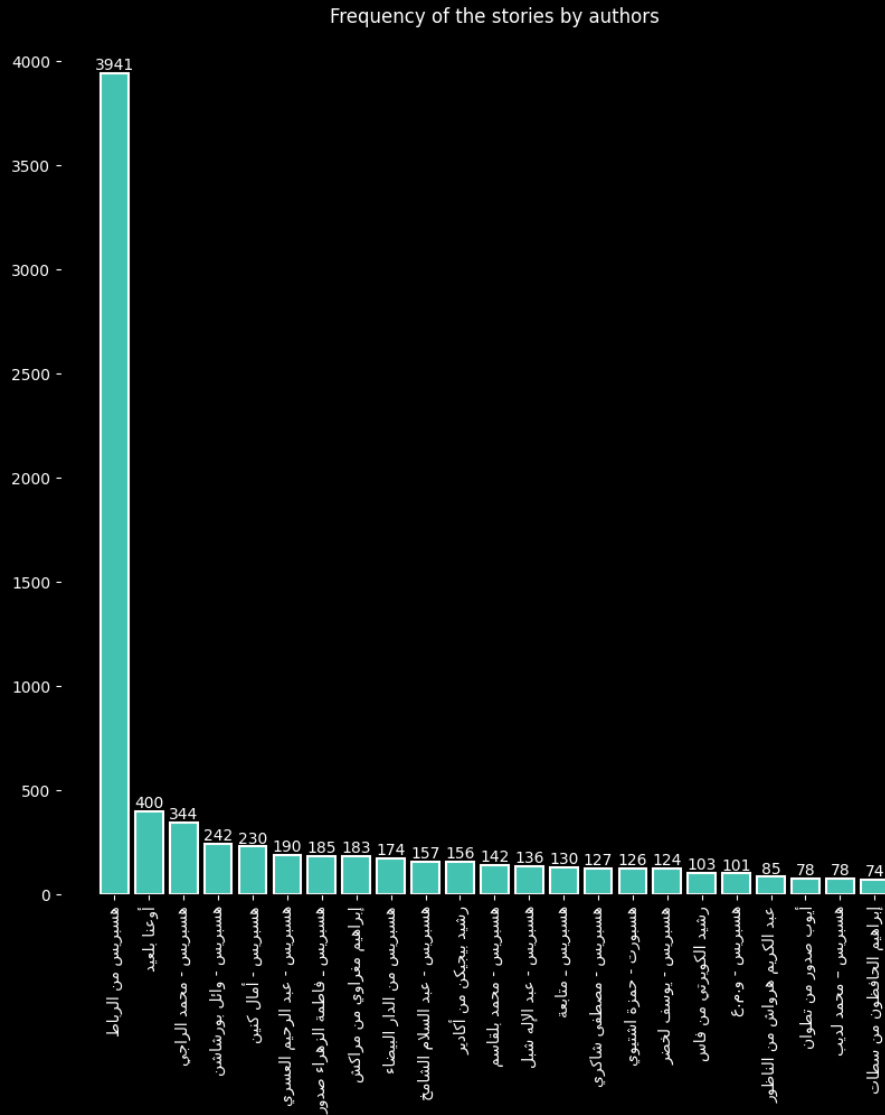


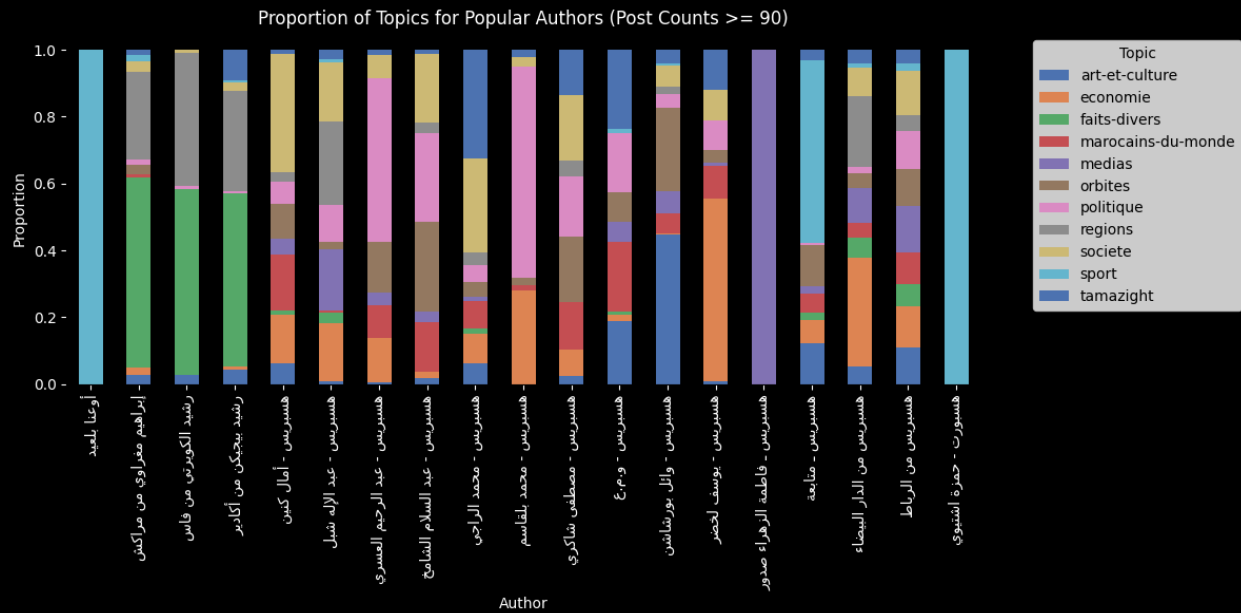
Hespress EDA

Using Hespress Dataset mainly the Stories dataset, we can note some of the visual insights that we found:

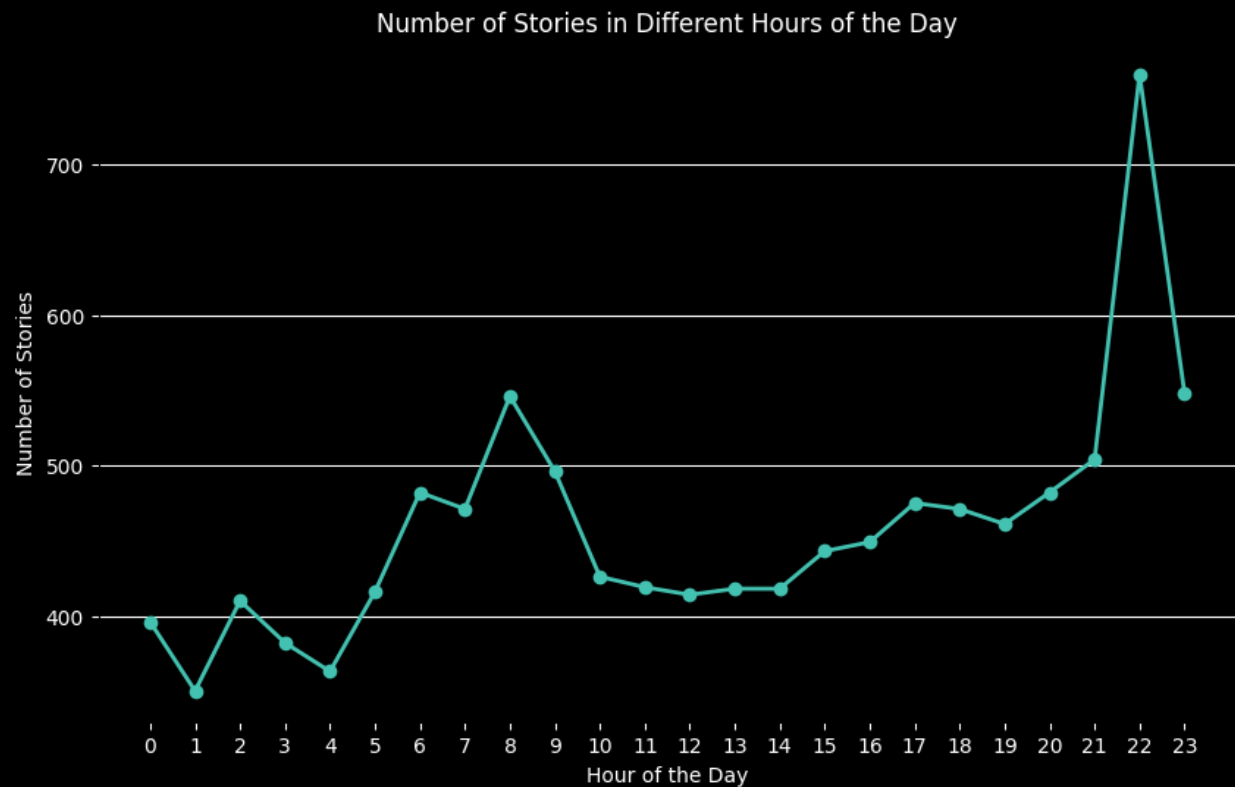
◆ Authors:



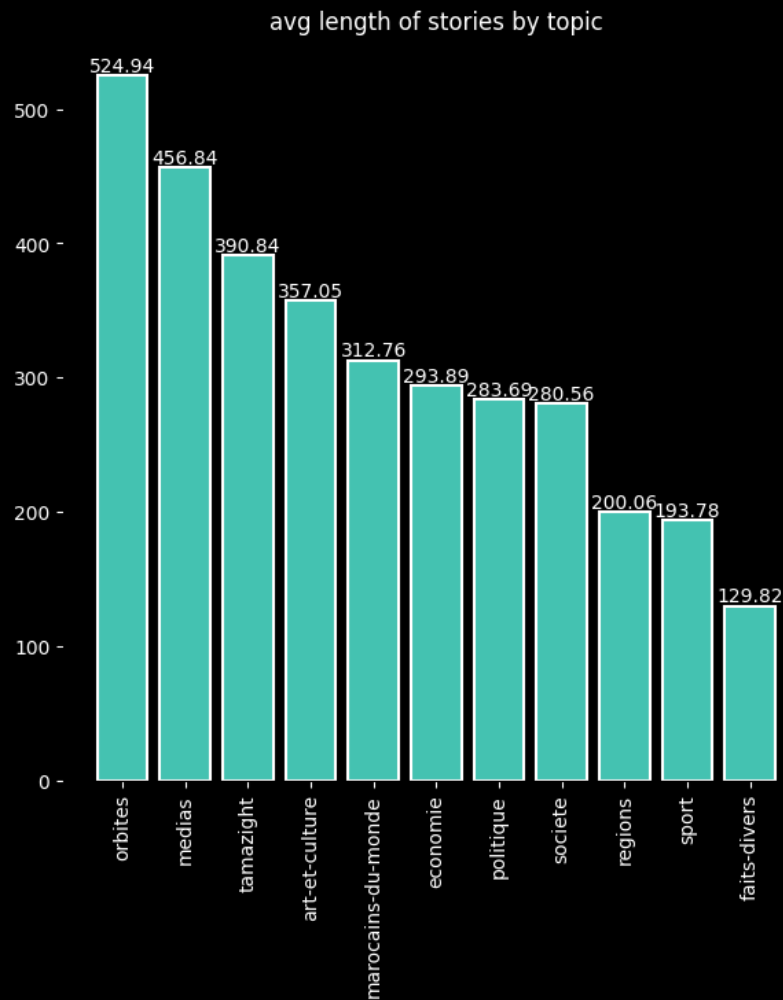
◆ Topic vs Authors :



◆ Date vs Stories published:



◆ Topics avg lengths:



Insights:

◆ Most stories are written by “هسبريس من الرباط”

◆ During the days hours, the peek of posting a story is at 10:00 pm

◆ Some authors are more concerned with a single topic such as those authors who write about “sports” or “faits-divers”

◆ There seems to be distinct variations in the lengths of the stories written in different topics. “Orbits” have the longest stories while “faits-divers” are the shortest

Note:

♦ **Most plots that include the authors have been thresholded with visually plotting only the authors with at least ~90 stories**

♦ **Many more potential plots could be exploited such as those related to the stories NER , word cloud , word frequency analysis or time series analysis .. etc.**