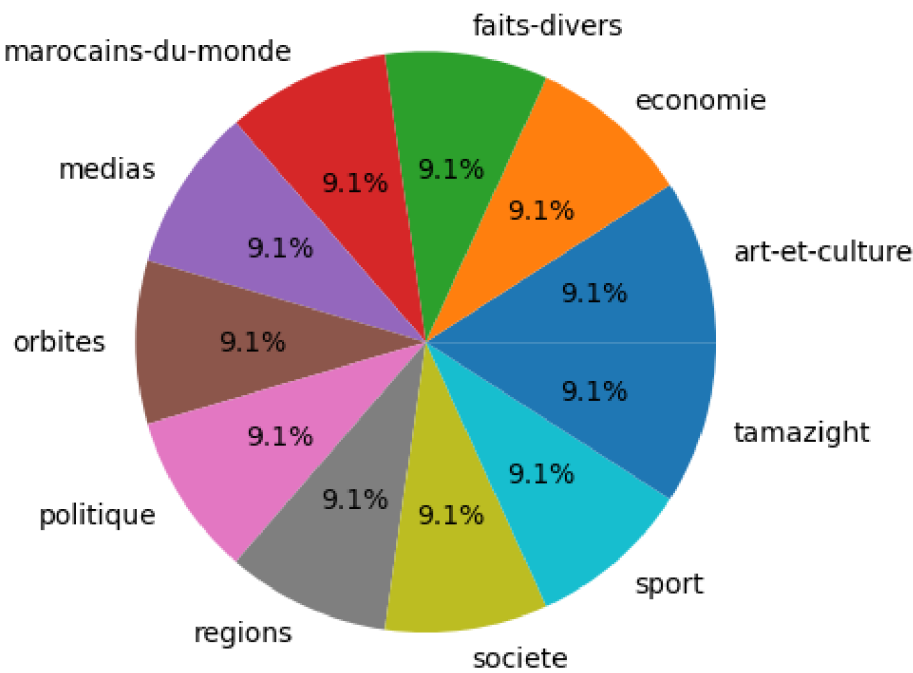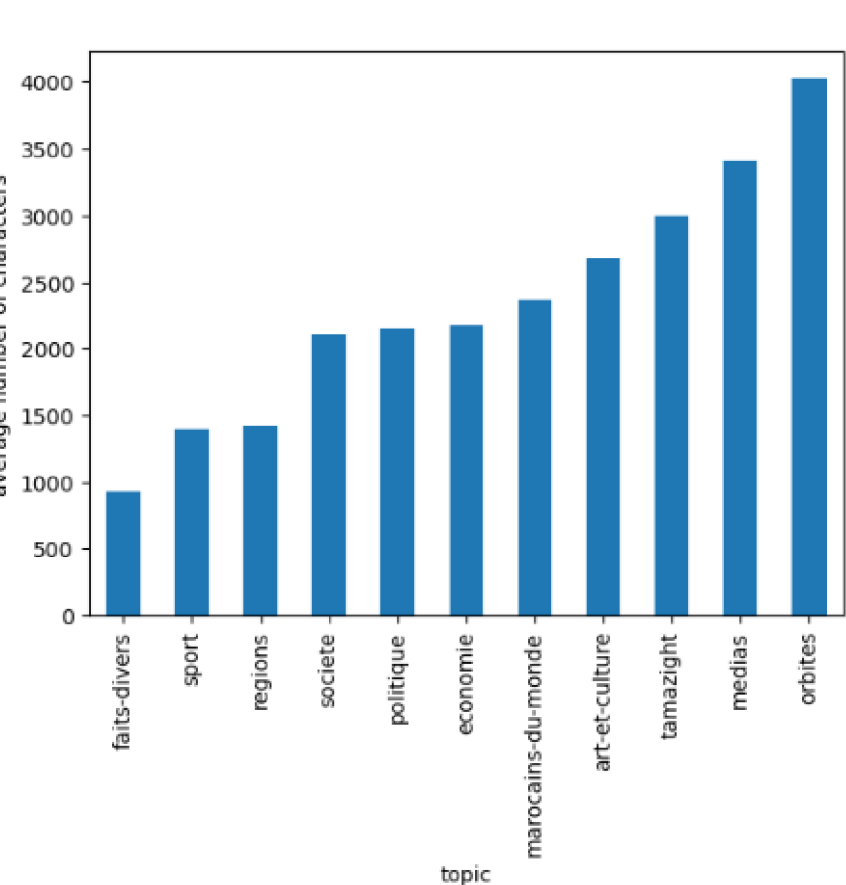# EXPLORING STORIES DATASET

Fig 1: Distributions of stories anda authors for each topic

DATASET IS BALANCED

As in fig 1, there are 11 different topics of stories, each one has the same number of stories. Even the distributions of authors across the topics are distributed equally.

Fig 2: the average number of characters in a story of each topic

topic	max	mean	min
faits-divers	5933	928.050	256
sport	10388	1398.471	151
regions	6962	1422.024	274
societe	17113	2100.177	358
politique	12555	2150.037	225
economie	12354	2177.350	289
marocains-du-monde	16629	2373.735	256
art-et-culture	33330	2675.401	256
tamazight	34371	3000.346	285
medias	14709	3413.936	248
orbites	33671	4026.461	134

Table 1: distributions of characters for each topic

STORIES LENGTH

each topic has a significant difference in the number of characters in each story as in Table 1.
the average number of characters is increasing for each topic as fig.2

Also. words number for each topic is on average increasing as in fig 3 with orbites as the highest topic with the most words on average for each story as in Table 2

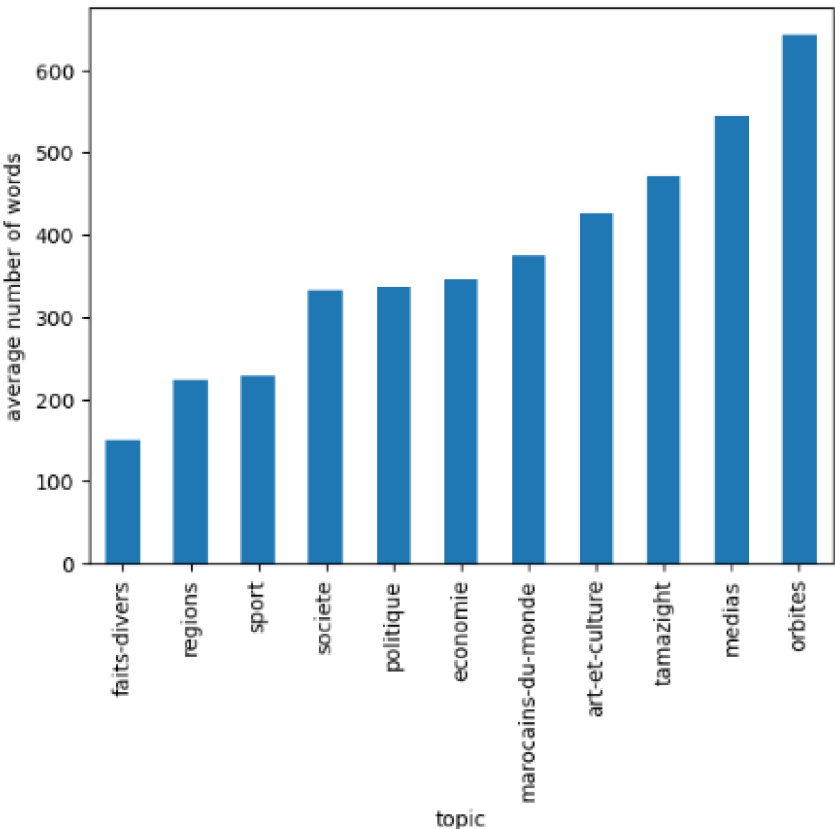| topic	max	mean	min
faits-divers	981	149.244	40
regions	1139	223.271	40
sport	1760	227.919	25
societe	2598	332.185	54
politique	2019	336.223	39
economie	1853	345.076	47
marocains-du-monde	2709	374.791	40
art-et-culture	5920	426.650	40
tamazight	5820	471.819	43
medias	2475	545.222	40
orbites	5720	643.044	24

Table 2: distributions of words for each topic

Fig 3: the average number of words in a story of each topic